

Tilburg University

Prosodic realizations of text structure

den Ouden, H.

Publication date:
2004

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
den Ouden, H. (2004). *Prosodic realizations of text structure*. In eigen beheer.

General rights

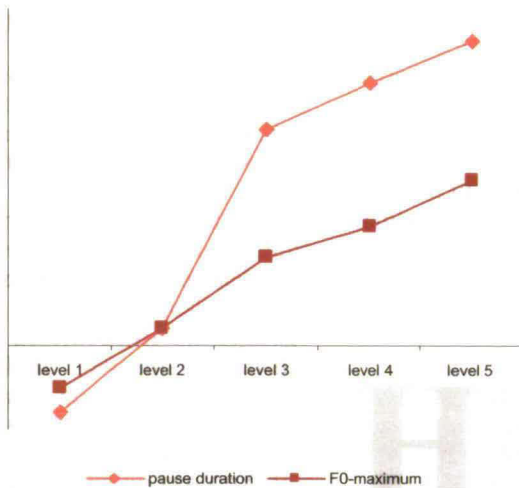
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

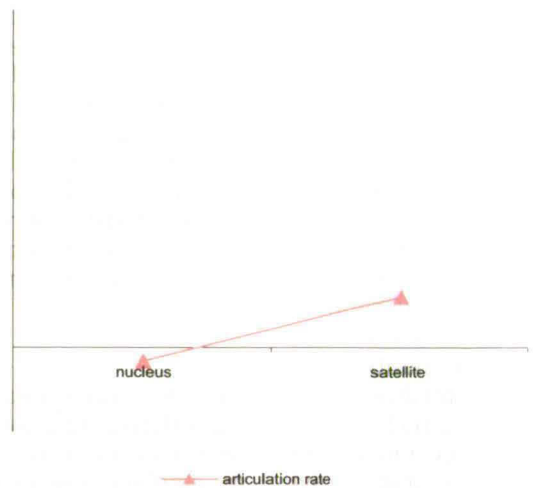
Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

hierarchy and prosody



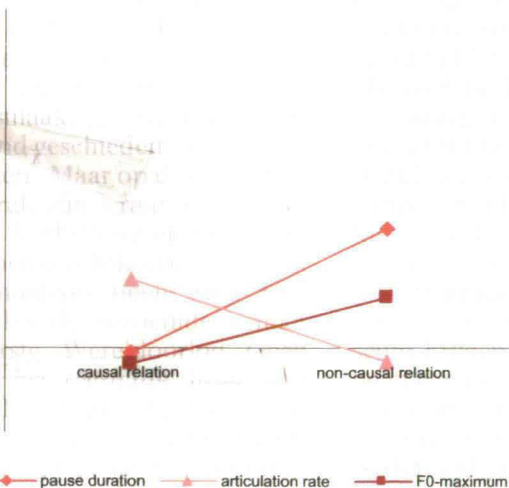
nuclearity and prosody



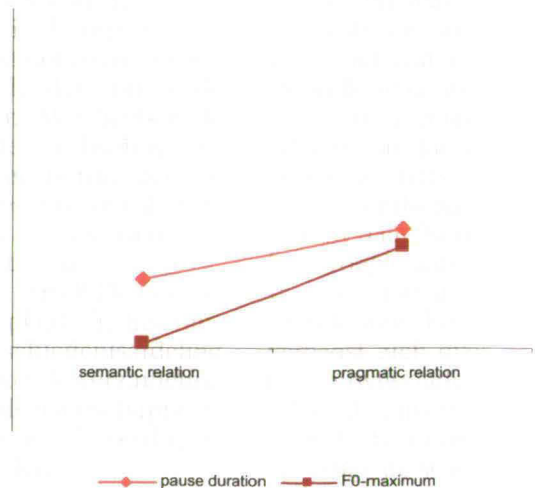
Prosodic realizations of text structure

Hanny den Ouden

causality and prosody



semanticity and prosody



Prosodic realizations of text structure

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Tilburg,
op gezag van de rector magnificus,
prof. dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen
ten overstaan van een
door het college voor promoties
aangewezen commissie
in de aula van de Universiteit
op maandag 20 december 2004
om 16.15 uur

door

Johanna Neeltje den Ouden

geboren op 22 maart 1964 te Hendrik Ido Ambacht



Promotor: Prof. dr. L.G.M. Noordman

Copromotor: Dr. J.M.B. Terken

The work in this thesis has been carried out under the auspices of the J.F. Schouten School for User-System Interaction Research, Department of Technology Management, Technische Universiteit Eindhoven, and of the Center for Language Studies, University of Tilburg, University of Nijmegen, and Max Planck Institute for Psycholinguistics. It was funded by SOBU (Cooperation of Brabant Universities).

Prosodic realizations of text structure

(met een samenvatting in het Nederlands)

Hanny den Ouden



Ouden, J.N. (Hanny) den

Prosodic realizations of text structure

Johanna Neeltje (Hanny) den Ouden

Proefschrift Universiteit van Tilburg.

Met lit. opg. - Met een samenvatting in het Nederlands.

ISBN 90-9018929-7

Trefw.: tekstanalyse, tekststructuur, tekstproductie, prosodie

Nederlandse titel: Prosodische realisaties van tekststructuur

© 2004 Hanny den Ouden

Cover designed by Otto Meijer

Cover realized by Bart Campman

Printed by Ipskamp, Enschede

Voor Marit, Tijmen en Nina
om verwonderd te blijven

Dankwoord

Dit proefschrift is tot stand gekomen in een gelukkige periode in mijn leven. Het onderzoek was inspirerend om te doen, het schrijven van het proefschrift een ware uitdaging voor mijn uithoudingsvermogen. In dezelfde tijd groeiden onze drie kleine kinderen op tot mooie kleine mensen. Ik heb genoten van de afwisseling tussen de levendigheid van ons gezin en de bedachtzaamheid die nodig was om onderzoek te doen. Bij dit alles wist ik me geruggesteund door een grote kring van mensen.

Ik bedank Leo Noordman en Jacques Terken, mijn promotores, voor hun enthousiasme en hun betrokkenheid bij mijn promotieproject. De adviezen en commentaren die ik van hen kreeg tijdens, en soms daags na, onze vele driekoppige besprekingen hebben de inhoud van dit proefschrift enorm verbeterd. Zo lang en zo intensief begeleid worden is een ervaring die voor mij met niets anders vergelijkbaar is. Vanuit de eigen aardigheden van drie mensen worden in zo'n proces zoveel kennis en ervaring overgedragen, ideeën bedacht en uitgewerkt, dat ik het, terugblikkend, niet anders kan zien dan als iets moois. Ik weet zeker dat ik later in mijn werk vaak aan uitspraken van Leo en Jacques herinnerd zal worden en dat zal me plezier doen.

Carel van Wijk, mijn informele begeleider, bedank ik voor het vertrouwen dat hij in mij stelde. Met een goede mengeling van waardering en kritische zin heeft hij me dikwijls over hobbels heen geholpen; het gebabbel, de grapjes, het op elkaar ingespeeld raken, dat alles maakte onze samenwerking bijzonder en plezierig. Ik zie ernaar uit om de vele ideeën die we in de voorbije jaren over tekstproductie en prosodie hebben gehad samen uit te kunnen werken.

Ik heb het als een voorrecht beschouwd om op twee plaatsen te mogen werken. Als AiO was ik tegelijkertijd verbonden aan het voormalige Instituut voor Perceptie Onderzoek aan de Universiteit van Eindhoven en aan de letterenfaculteit van de Universiteit van Tilburg. Zo leerde ik twee totaal verschillende onderzoeksculturen kennen en twee andere organisatievormen; ik had op twee plaatsen aardige en inspirerende collega's en op twee plaatsen onmisbare kamergenoten. Met Marc Swerts deelde ik de kamer in Eindhoven, met Birgit Bekker en later Leonoor Oversteegen de kamer in Tilburg. Ik bedank hen alle drie voor de gezelligheid, de uitwisseling van ervaringen op onderzoeks- en onderwijsgebied, de stimulans die daarvan uitging, en het delen van het wel en wee van iedere dag. Ook was het fijn op beide plaatsen collega-AiO's te hebben met wie ik ongeveer gelijke tred hield: met Anja Arts en Olga van Herwijnen zette ik de eerste schreden op de weg van cursussen, presentaties en buitenlandse reizen.

Sommige mensen waren wat meer op afstand betrokken bij mijn project, maar ik kon op hen altijd een beroep doen. Wilbert Spooren was zo iemand. Als ik kritisch commentaar nodig had juist van iemand aan de zijlijn, dan was hij de aangewezen persoon. Meermalen behoeftte hij me voor uitglidders. Bill Mann, die ik ontmoette op een conferentie in Lyon in 2000, was ook zo iemand. Het doet me verdriet dat hij stierf enkele maanden voordat ik dit proefschrift afrondde.

Dankwoord

Ook ben ik dank verschuldigd aan de collega's, studenten, familieleden, vrienden en kennissen die optraden als tekstanalisten, beoordelaars van zinnen en spraakuitingen, of sprekers in mijn onderzoeken. Speciale dank gaat uit naar Annelien Scheele en Peter Wisse die het onderzoek uitvoerden dat in Hoofdstuk 6 gerapporteerd staat. Ik dank Lauraine Sinay, Anneke Smits en Patricia Goldrick van harte, omdat zij mij met allerlei hand-en-span diensten terzijde stonden, me hielpen met de uiteindelijke afwerking van het manuscript en mijn Engels corrigeerden. Het is fantastisch om in de laatste hectische fase zulke praktische dingen aan zulke capabele mensen toe te kunnen vertrouwen.

Onze kring van familie en schoonfamilie, met name mijn schoonouders, bedank ik voor hun aanmoediging en belangstelling voor mijn werk. Ik betreur het dat mijn ouders de verdediging van mijn proefschrift niet meer mee kunnen maken, maar het vervult me met dankbaarheid dat ik mij omringd weet door mijn zeven oudere broers, zussen en hun partners. Coen Goossens, Ellen Hoeckx, Deborah Hurks, Ton Jansens, Jan Pekelder, Sylvie Adnet, Herlinde van Dijck, Jurriaan Balke, Hannelore Lubinski, Lia van de Laar, Wilma Wijnen en Jan Schellekens dank ik van harte voor wat zij me gaven: hun vriendschap, aanmoediging, belangstelling en steun.

Ik draag deze dissertatie op aan onze kinderen, Marit, Tijmen en Nina, omdat ik ze toewens dat zij zich gedurende hun hele leven vragen blijven stellen over de wereld om hen heen, en de schijnbaar eenvoudige dingen in het leven niet voor vanzelfsprekend zullen aannemen.

Ad van Liere, mijn levenspartner, dank ik het meest, voor het echt samen delen van de zorg voor ons gezin, voor zijn onevenredig grote aandeel daarin tijdens de laatste fase van dit proefschrift, voor de ruimte die hij me gaf, zijn goede raad op velerlei gebied, zijn geduld, zijn humor en liefde.

Hanny den Ouden,
november 2004

Contents

1	Introduction	1
1.1	Research topic	3
1.2	Research questions	7
2	Reliability of text structure analyses	11
2.1	Introduction	13
2.2	Text structure analyses	15
2.2.1	Intuition-based procedures	15
2.2.2	Theory-based procedures	16
2.3	Method	23
2.3.1	Text material	23
2.3.2	Procedures	23
2.3.2.1	Free task	23
2.3.2.2	Restricted task	24
2.3.2.3	Intention Based Analysis	24
2.3.2.4	Rhetorical Structure Theory	24
2.4	Results	25
2.5	Conclusion and discussion	27
3	Reliability of pitch-range measurements	29
3.1	Introduction	31
3.2	Pitch-range measurements	32
3.3.1	Judges	34
3.3.2	Speech material	34
3.3.3	Procedure	34
3.4	Results	35
3.4.1	Effect of length on pitch-range measurements	35
3.4.2	Agreement between human pitch-range measurements	36
3.4.2.1	Relative agreement	37
3.4.2.2	Absolute agreement	38
3.4.3	Agreement between human judges and automatic measurements	40
3.4.4	Relevance of F0-maximum as characterization of pitch range	41
3.5	Conclusion and discussion	42

4	Prosody of hierarchy: An exploration	43
4.1	Introduction	45
4.2	Method	46
4.2.1	Text material	46
4.2.2	Three procedures for scoring hierarchical level	46
4.2.3	Speech material	51
4.2.4	Two approaches for relating text structure and prosody	52
4.3	Results	55
4.3.1	Effect of syntactic status on prosody	55
4.3.2	Relative approach	57
4.3.2.1	Procedure of linear adjacency	57
4.3.2.1.1	Top-down scoring	58
4.3.2.1.2	Bottom-up scoring	59
4.3.2.1.3	Symmetrical scoring	60
4.3.2.2	Procedure of hierarchical adjacency	61
4.3.2.3	Summary of results	64
4.3.3	Absolute approach	64
4.3.3.1	Top-down scoring	65
4.3.3.2	Bottom-up scoring	66
4.3.3.3	Symmetrical scoring	68
4.3.3.4	Summary of results	69
4.4	Conclusion and discussion	71
5	Prosody of hierarchy, nuclearity, and rhetorical relations: A corpus-based study	73
5.1	Introduction	75
5.2	Method	75
5.2.1	Text material	75
5.2.2	Text-structural characteristics	76
5.2.2.1	Hierarchy	76
5.2.2.2	Nuclearity	76
5.2.2.3	Rhetorical relations	76
5.2.3	Procedure	80
5.2.4	Speech material	80
5.3	Results	82
5.3.1	Effect of syntactic class on prosody	82
5.3.2	Effect of hierarchy on prosody	83
5.3.2.1	Relative approach	83
5.3.2.2	Absolute approach	86
5.3.3	Effect of nuclearity on prosody	89
5.3.4	Effects of rhetorical relations on prosody	90
5.3.4.1	Causal and non-causal relations	90
5.3.4.2	Semantic and pragmatic relations	91
5.4	Conclusion and discussion	93

6 Prosody of causal and non-causal, and of semantic and pragmatic relations:	
Two experiments	99
6.1 Introduction	101
6.2 Experiment 1: Prosody of causal and non-causal relations	101
6.2.1 Pretest: Construction and selection of text material	102
6.2.1.1 Text material	102
6.2.1.2 Judges	103
6.2.1.3 Procedure	103
6.2.1.4 Results	104
6.2.2 Main study: Prosodic realization of causal and non-causal relations	104
6.2.2.1 Speakers	104
6.2.2.2 Procedure	104
6.2.2.3 Speech material	105
6.2.2.4 Results	105
6.2.2.5 Discussion and conclusion	108
6.3 Experiment 2: Prosody of semantic and pragmatic relations	110
6.3.1 Pretest: Construction and selection of text material	110
6.3.1.1 Text material	110
6.3.1.2 Judges	113
6.3.1.3 Procedure	113
6.3.1.4 Results	114
6.3.2 Main study: Prosodic realization of semantic and pragmatic relations	115
6.3.2.1 Speakers	115
6.3.2.2 Procedure	115
6.3.2.3 Speech material	116
6.3.2.4 Results	116
6.3.2.5 Discussion and conclusion	119
6.4 Conclusion and discussion	121
7 Discussion	123
7.1 Conclusions	125
7.1.1 Analyses of text structure	125
7.1.2 Measurements of prosody	126
7.1.3 The relation between text structure and prosody	126
7.2 Implications for text-to-speech systems	128
7.3 Beyond the limitations of this research	129
References	133

Contents

Appendix A	Original Dutch text of the sample text used in Chapter 2	139
Appendix B	Original Dutch text of the sample text used in Chapter 4	140
Appendix C	Original Dutch text of the sample text used in Chapter 5	141
Appendix D	Instruction pretest: causality test	142
Appendix E	Instruction pretest: plausibility test	143
Appendix F	Instruction pretest: semanticality test	144
Appendix G	Texts used in the experiment on causal and non-causal relations	145
Appendix H	Texts used in the experiment on semantic and pragmatic relations . . .	149
Summary in Dutch		153
Curriculum Vitae		159

1

Introduction

1.1 Research topic

The focus of this dissertation is on the relation between text structure and prosody. Text structure pertains to the organization of a text. A text is a collection of sentences¹ that cohere in some way: each sentence is related to another sentence or to a group of sentences. The organization of a text and the coherence between the sentences can be represented as a hierarchical structure. Most theories in the field of discourse studies represent hierarchical structures of texts as fully connected trees with branches, the end nodes of which are the individual sentences of the text. In such a hierarchical representation, a central sentence corresponds with a higher position in the hierarchy than a less central one. Psycholinguistic research has demonstrated that hierarchical representations of text structures have cognitive plausibility; for example, it has been shown that sentences at high positions are better recalled than sentences at low positions. This phenomenon is called the 'levels effect' (Singer, 1990: 40).

Prosody pertains to the suprasegmental aspects of speech, i.e., characteristics beyond the level of the individual speech sounds of vowels and consonants. Prosody is made up of a heterogeneous set of features which contains at least pausing, speech rate, phrasing, intonation, rhythm, accentuation, and loudness. Most research on prosody has focused on the prosody of sentences. The prosody of sentences has been described in detail in terms of accentuation patterns and intonation contours, for Dutch, for example, by 't Hart, Collier, & Cohen (1990). Text prosody is concerned with prosodic characteristics beyond the level of sentences. The prosody of texts, however, has been investigated far less extensively. Sentences in isolation differ from sentences in the context of texts. It can be perceived when people talk that prosodic features do not precisely correspond with the domain of a single sentence: often, prosody seems to run over sentences (Swerts & Geluykens, 1993). In the field of speech technology, also, it has been found that the prosody of texts is not merely the sum of the prosody of sentences. For example, although sentences generated by a computer may sound quite natural when heard in isolation, they do not sound as natural when they are simply concatenated and combined in a text (Silverman, 1987, Terken, 1993). Therefore, the prosody of text seems to require components to be added to the rules governing the prosody of sentences. A further delineation of these components constitutes the topic of this dissertation.

The relation between text structure and prosody might be considered analogous to the relation between text structure and typography in written texts. Prosody in spoken texts may function as typography in written texts. The writer of a text applies many typographical means, such as punctuation marks, capital letters, italics and bold, blank lines, indentation, footnotes, and a division into sections and paragraphs. They help a writer to convey the structure of the text as he or she conceptualized it. These markers can help a reader to recover the structure, and, therefore, to understand the message more easily. Analogously, a speaker may apply various prosodic means such as variation in pause duration, articulation rate, and intonation to convey the structure underlying the text, and consequently to help a listener understand the message more easily.

¹ The term 'sentence' is loosely used in this chapter. In a strict sense, 'clause' is meant. Within the framework of text analysis as applied in the following chapters, the clause is referred to using the term 'segment'.

Earlier research on text prosody concentrated on the prosodic marking of two textual levels: sentences and paragraphs. Prosodic differences were demonstrated at paragraph boundaries and boundaries between sentences within paragraphs. Paragraph boundaries are associated with longer pauses than boundaries between sentences within paragraphs (Lehiste, 1979; Silverman, 1987). A lowering of successive fundamental frequency peaks and valleys over sentences within a paragraph was also observed (Bruce, 1982; Brown & Yule, 1983; Thorsen, 1985; Sluijter & Terken, 1993). The final sentences of paragraphs, and parentheticals were found to be articulated with lower pitch range and faster speech rate than sentences at other locations in texts (Brubaker, 1972; Lehiste, 1975; Brown, Currie, & Kenworthy, 1980; Grosz & Hirschberg, 1992; Koopmans & Van Donzel, 1996). These studies showed that prosody has a function in the marking of the coherence of sequences of sentences: some sentences in a text are more strongly connected to each other than other sentences, and this difference is marked using prosodic means.

Later research on text prosody distinguished more than two textual levels. It concentrated on various kinds of boundaries between text units: boundaries between text units may differ in 'weight'. Prosody associated with 'stronger' boundaries differs from prosody associated with 'weaker' boundaries (Swerts, 1997; Schilperoord, 1996; Hirschberg & Nakatani, 1996; Noordman, Dassen, Terken, & Swerts, 1999; Smith & Hogan, 2001). These studies showed that the durations of pauses and the heights of fundamental frequency gradually decrease as the boundaries between text units become weaker.

This dissertation brings together the concepts of the hierarchical structure of a text and its prosodic marking when speakers articulate the text. High-level boundaries in a text structure are considered strong boundaries, whereas low-level boundaries in a text structure are considered weak boundaries. The text units separated by high-level boundaries are more loosely connected than the text units separated by low-level boundaries. In the same way as higher-level sentences are better recalled than lower-level sentences, we expect that speakers might mark higher-level boundaries using stronger prosodic cues than those used for lower-level boundaries. Hierarchical representations of texts also provide information about the nuclearity of sentences, and the specific ways in which sentences are related. In addition to examining the relation between hierarchy and prosody, it is investigated whether the nuclearity of segments and rhetorical relations are reflected in prosody.

For a clear understanding of the approach adopted in this dissertation, some methodological points have to be discussed first: the use of natural texts, the use of prepared speech, and the selection of prosodic features.

In line with the trend towards corpus research that has been inspired by language and speech technology, the focus of this dissertation is explicitly on natural texts. Unlike most previous research on the relation between text structure and prosody, this dissertation aims to study text prosody *in vivo*, i.e., in speech materials that were not created specifically for the purpose of the research. In earlier research on the prosodic realization of aspects of text structure, constructed texts with predetermined paragraph boundaries were often used (for example, Brubaker, 1972; Bruce, 1982; Thorsen, 1985; Lehiste, 1975; Sluijter & Terken, 1993; Silverman, 1987; Noordman et al., 1999). Others used spontaneous speech that was tightly constrained by experimentally eliciting texts in such a way that they were easy to divide in separate information units (for

example, Terken, 1984; Swerts & Collier, 1992; Geluykens & Swerts, 1994; Caspers, 2000, Mushin, Sterling, Fletcher, & Wales, 2003). The studies reported in Chapters 2 to 4 make use of speech materials that were broadcast on Dutch radio; the study described in Chapter 5 makes use of news reports published in a Dutch national quality newspaper (although for reasons explained in Chapter 5 the texts were read aloud specifically for the purpose of this study). Only in the study described in Chapter 6 constructed texts are used, because specific hypotheses about the relation between text structure and prosody had to be tested under controlled circumstances. In all these studies, it is explored whether natural texts can be used to determine such aspects as the reliability of text structure analyses or the robustness of the relation between text structure and prosody.

Parallel with the distinction between natural and constructed texts is the distinction between the corpus-based approach and experimental design. The texts used in the study reported in Chapter 4 are a small corpus to explore different procedures for quantifying scores for hierarchical structure and to explore ways to relate the scores for text structure to the scores for prosody. The texts used in the study reported in Chapter 5 form a larger text corpus, used in the investigation of the relations between hierarchy, nuclearity, and rhetorical relations, on the one hand, and prosody, on the other hand. The advantage of the use of corpora is that the relation between text structure and prosody can be sought in textual contexts affected by various factors: the robustness of the relation can be demonstrated. The disadvantage is that, in addition to text-structural features, they contain confounding variables which cannot be controlled for. The experimental approach in the study described in Chapter 6 adjusts for this general shortcoming of a corpus-based approach. It made it possible to assign prosodic findings unambiguously to specific text-structural aspects.

The objective of this dissertation required the use of spoken texts of such a length that hierarchical structures of sufficient depth could be obtained. Long monologues had to be made available, for example, by having speakers tell a story spontaneously, or by having readers read aloud written texts. Spontaneous and read-aloud speech differ in many ways (Johns-Lewis, 1986; Ayers, 1992; Blaauw, 1992). Spontaneous speech is not prepared in advance: while talking spontaneously, speakers have to think about what they are going to say and how they are going to say it, causing hesitations, silences, repetitions, self-repairs, and so forth. For the demonstration of the relation between text structure and prosody, in spontaneous speech it cannot be determined unambiguously whether prosodic features are the result of the planning activity during speaking or the structuring of the text. This holds for spontaneous monologues spoken in isolation, or in interaction. Extended monologues can be elicited in isolation by asking speakers to produce a text with a particular text structure, for example, a description of a structured object like a house (Terken, 1984, Swerts & Geluykens, 1994) or a route on a map (Caspers, 2000, Mushin et al., 2003), but still this kind of text material does not show unambiguously whether the prosodic features are caused by the structuring or the planning activity of the speaker. Spontaneous speech in interaction is also affected by planning activity during speaking, but, also, by the interaction process with the listeners. Monologues in a conversational situation are influenced by non-verbal signals of the listener(s). Even though the listeners may not be physically present, for example, in multimedia situations, like newscasts, the monologues are

influenced by the interaction with the television pictures (Oviatt & Cohen, 1991). To separate the effects of the structuring function of prosody from those of the planning function and the interaction function of prosody, in this dissertation, neither spontaneous speech in isolation nor spontaneous speech in interaction could be used.

Long read-aloud texts were used. Speakers knew beforehand what they were going to say since the written text was given. They did not have to decide or plan what they are going to say. Read-aloud speech can be prepared and unprepared. In an unprepared reading-aloud task, speakers start reading aloud straightaway; in a prepared reading-aloud task, speakers read the written text several times and only then read it aloud. In order to know how they are going to say it, speakers need to prepare the text. To demonstrate that speakers realize text structure prosodically, awareness of the structure of the text is a prerequisite. This awareness of text structure is probably optimal for the author of the text. Therefore, for a part of the research reported in this dissertation, long texts were read by speakers who were the authors. For another part of the research, long texts were read aloud by speakers who were not the authors, but they were asked to prepare the texts conscientiously and extensively in order to make them aware of the structure of the texts.

We concentrate on three prosodic features: pause duration between sentences, pitch range, and the articulation rate of sentences. They are the primary relevant prosodic features for demonstrating the relation between hierarchical structure and prosody. If the relation between hierarchy and prosody is established for these primary prosodic features, prosodic features of secondary importance would have to be looked for as well. These features would be, for example, preboundary lengthening, final lengthening, peak displacement, filled pauses, contour differences, and intensity, since these features were found to be related with boundary strength and topical structure (Ladd, 1988; Swerts, 1993; Swerts, Bouwhuis, & Collier, 1996; Wichman, House, & Rietveld, 1997; Swerts, 1997; Smith & Hogan, 2001). Intensity or loudness has probably a relation with hierarchical structure. It is complicated to measure, however, because the angle of the speaker in relation to the microphone, and his or her distance from it, must be controlled for. A considerable number of other prosodic features could have been measured as well, and they would possibly have shown a relationship with some aspects of text structure (Batliner, Buckow, Huber, Warnke, Nöth & Niemann, 2001). We are primarily interested, however, in the prosodic realization of the hierarchical aspect of text structure. Abundant evidence is available that pause duration, pitch range, and articulation rate are sensitive to the positions of utterances in a text structure (Hirschberg & Nakatani, 1996; Swerts, 1995, 1997). In this dissertation, those prosodic parameters are selected which were considered most likely to mark hierarchical aspects of text structure.

The results of this dissertation are based solely on acoustical analyses of the speech material, and not on perceptual analyses. This is because it is our objective to provide empirical evidence for the relation between text structure and prosody rather than to account explicitly for the relation in a theoretical model of language use and language interaction in which both production and perception play a role; see, for example Levelt (1989) or Clark (1996). The perspective of the

dissertation is entirely production-oriented, although the findings for the prosodic marking of text structure can be explained in a perception perspective as well.

The aim of this research is to contribute to the theory of human text production and to the improvement of automatic text-to-speech systems. One contribution to the theory of human text production is that the prosodic patterns human speakers realize to structure texts that are found in this study support the psychological plausibility of text-structural notions. People who work on automatic text-to-speech systems may be helped by the prosodic correlates of text-structural components found in this study.

1.2 Research questions

The research objectives of this dissertation are twofold. The first objective is to contribute to the theoretical modeling of human text production. A close relation between text structure and prosody would add to our knowledge of both the planning and formulating processes of speakers: apparently, speakers are aware of the rhetorical organization of their messages and try to convey it to their listeners. Theories of human text production will be more complete when the prosodic patterns speakers use when they convey text-structural information are known.

The second objective is to contribute to the improvement of automatically generated texts. If it can be shown that prosodic features coincide systematically with text-structural aspects, text-to-speech systems may benefit from this by explicitly keeping track of the structure of the text under construction and by adjusting prosodic parameters in accordance with this structure. Texts will sound more natural than they do without the application of text prosody.

In the realization of these objectives, they are phrased in more specific terms. The aims of the studies reported in this dissertation are to contribute to the clarification of the relation between text structure and prosody by providing empirical evidence for three research domains: first, the reliability and relevance of procedures for assigning text structure; second, the reliability and relevance of physical measurements of prosody; and third, the actual relation between texts and their prosody. The first two domains of research concern steps needed to prepare for the third, which addresses the research topic itself.

Two lines of research, both with a long-standing tradition, come together in the research on the relation between text structure and prosody. Nevertheless, standard operationalizations of text-structural and prosodic characteristics needed for the evaluation of their relation were lacking. Before the two lines could come together, two preparatory steps had to be taken on each of the research topics separately. First, both in the field of prosody and in the field of text structure, the elements of observation had to be selected. Second, for prosody and text structure separately, the observations had to be transformed into scores. Figure 1.1 depicts the general set-up of this dissertation in terms of the three steps: observations, scores, and evaluation. The order in the figure corresponds with the order of the studies reported in this dissertation. The research questions of this dissertation are introduced using Figure 1.1 as a framework.

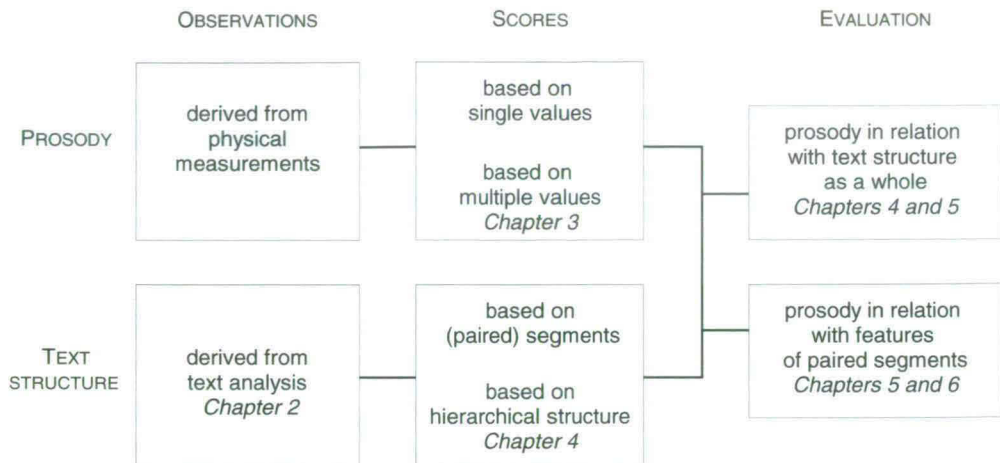


Figure 1.1 Schematic representation of the research activities and where they are reported

The general questions for both the field of prosody and the field of text structure are: how to get observations (left side of Figure 1.1), how to transform observations into scores (middle part), and how to evaluate the relation between both kinds of scores (right side).

On the left side of Figure 1.1, the observations of interest for both prosody and text structure are shown. For prosody, this step did not pose a great problem: observations were derived from physical measurements in the speech signal using technical equipment. The prosodic features of the speech signal relevant to text prosody were pause durations between sentences, the speed of speaking, and pitch range. In this dissertation, the physical measurements provide information on pause duration, fundamental frequency, and articulation rate.

For text structure, however, such automated registrations were not possible. The structure of a text could only be ‘observed’ using a hand-made text analysis. Text analyses can be made on the basis of theoretical accounts (such as Thorndyke, 1977; Mann & Thompson, 1988; Grosz & Sidner, 1986; Sanders & Van Wijk, 1996) and more intuitive accounts (such as Rotondo, 1984; Sluijter & Terken, 1993; Swerts, 1997). In Chapter 2, two intuition-based and two theory-based procedures for assigning multilayered hierarchical structures to the same four texts are described. In the intuitive procedure of annotating text structure, naive subjects indicated major boundaries in the texts using a more and a less restricted variant of the procedure. Scores for the hierarchical levels of the text structure were obtained by counting the number of subjects who annotated a boundary as a major boundary. In the theory-based procedures, three experts applied the theory proposed by Grosz & Sidner (1986) and six experts applied Rhetorical Structure Theory (Mann & Thompson, 1988). The object of study was the reliability of the procedures: do analysts come up with the same text structure analyses of a text when they apply a particular procedure independently of each other?

In the middle part of Figure 1.1, the transformation of observations into scores for both prosody and text structure is shown.

For prosody, the prosodic characteristics pause duration and articulation rate did not pose a problem, because they were based on ‘single’ values which were derived directly from observations registered automatically or indicated by hand. For pause duration, the duration of a stretch of silence in between stretches of speech was measured. For articulation rate, the number of phonemes or syllables in a given stretch of speech were counted. The transformation of pitch-range observations to scores was more problematic, because the pitch contour of a stretch of speech has ‘multiple’ values, i.e., a pitch contour consists of many pitch-range measurements during the articulation of the speech. To characterize the pitch range of a whole stretch of speech by a single score, in the study reported in Chapter 3, two ways of characterizing the pitch range of an utterance were examined, namely, using the highest peak of the contour (Lieberman & Pierrehumbert, 1984) and using the distance between two trend lines connecting the peaks and the valleys of the contour (‘t Hart, Collier & Cohen, 1990; Ladd, 1990). The object of study was the reliability of these characterizations: do analysts come up with the same estimations of the two pitch-range parameters when they apply the measurements independently of each other? In forty utterances, the highest peaks (F0-maxima) and the declination lines were determined by five trained phoneticians. It was investigated which way of characterizing pitch range was the most reliable and relevant one to apply.

For text structure, the scoring did not pose a problem for characteristics of (paired) segments, since these characteristics could be observed directly from the text or the text structure analysis. Characteristics of segments that can be observed directly are the syntactic status of segments (for example, whether they are main or subordinate sentences), the nuclearity of segments (whether a segment is a nucleus or a satellite), and the rhetorical relations between segments (for example, whether they are causally or non-causally related). The scoring of characteristics of the whole hierarchical structure was more problematic. The aim of the study described in Chapter 4 was twofold. First, several ways to give scores to the hierarchical levels of a text structure were explored. RST text structures are represented by tree-like figures consisting of branches with nodes at various levels. Three procedures to quantify these levels were investigated: top-down, bottom-up, and symmetrical procedures. Second, in preparation for the following studies, the relation between text structure and prosody was explored. Pause duration between successive sentences, the F0-maximum, and the articulation rate of the sentences were measured and related to the levels of the boundaries in the hierarchical structures of the texts. The levels could be scored on an interval scale and then related to their mean prosodic realizations, or the levels of two adjacent or superordinate boundaries could be scored ordinally in terms of ‘higher’ and ‘lower’ positions in the text structure and then related to their prosodic realizations.

On the right side of Figure 1.1, the relation between text structure and prosody is shown. Once the reliability of both prosodic and text-structural observations was established and relevant scores from these observations were obtained, the relation between text structure and prosody could be evaluated. It was evaluated in two ways in the core studies of this dissertation. Chapter 5 reports a corpus study of the prosody of sentences in relation to textual characteristics derived from text structures as a whole. Twenty news reports, read aloud by different speakers, were

analyzed using Rhetorical Structure Theory. RST provides a multilayered hierarchical structure of a text, it distinguishes nuclei and satellites within a text, and it identifies the rhetorical relations between the sentences in a text. Pause duration between adjacent sentences, the F0-maximum, and the articulation rate of the sentences were measured and related to the levels of the boundaries in the hierarchical structures of the texts, the nuclearity of the sentences, and the rhetorical relations between the sentences. The aim of the study was to examine how these text-structural characteristics are reflected by prosody.

It was difficult to test specific hypotheses using the natural text material. Therefore, two experiments were run on the prosodic realizations of causal and non-causal relations, and semantic and pragmatic relations. The experiments are reported in Chapter 6. Target sentences were constructed which were either causally or non-causally, or semantically or pragmatically, related to a preceding sentence. The target sentence and its preceding sentence were part of a short text. More than twenty speakers read these texts aloud. In the speech material, pause durations preceding and following the target sentences were measured, as were the F0-maximum, mean pitch range, and articulation rate of the target sentences. The prosodic characteristics of the target sentences in both conditions were compared. The questions were whether, under controlled circumstances, the prosody of causal relations differs from that of non-causal relations, and whether the prosody of semantic relations differs from that of pragmatic relations.

Chapter 7 gives the conclusions and explains the implications for further research.

2

Reliability of text structure analyses

2.1 Introduction¹

Text structure refers to the way texts are organized into paragraphs, sentences, and clauses, and to the relations between them. Paragraphs, sentences, and clauses may cohere in all kinds of ways; for example, a sentence or paragraph can be a reason, a cause, or an elaboration in relation to another sentence or paragraph. Some sentences or paragraphs can contain more important information than others. In a pair of paragraphs, one paragraph may express the central idea whereas the other paragraph may simply be an elaboration or clarification. Similarly, at the level of sentences within a paragraph, one sentence may be more important than the others as it expresses the crucial content within the paragraph, i.e., the content that one would expect to find in a summary of the text (Marcu, 1999). In a graphical hierarchical representation of text, an important paragraph or sentence will be given a higher position in the hierarchy than a less important one.

In the studies described in this dissertation, text structure analyses must meet three requirements. First, they have to be generally applicable to all kinds of texts. The procedures for analyzing texts may not put constraints on the domain and content, or type and length of the texts. Second, text structure analyses must offer the possibility of ascribing numeric scores to hierarchical levels. Therefore, the procedure for analyzing text structure has to make it possible to ‘weigh’ the importance of individual sentences in the text so that the positions of sentences in a hierarchical structure reflect their information value for the text as a whole. Text structures, therefore, are represented as tree-like structures in this research. Finally, text structures have to be analyzed reliably. When several observers analyze the hierarchical structure of a text, they have to give the same structure to it. This chapter is concerned with the reliability of text structure analyses.

Intuitive and theoretically motivated procedures can be used to analyze the hierarchical structure of a text. The use of intuitive procedures is common practice in prosodic research on texts, whereas theoretical procedures have been developed in text linguistics.

An intuitive procedure often applied in the field of prosody is that of asking subjects to judge text structure in texts, for instance, by indicating the locations of paragraph boundaries. These boundaries indicate the locations in the text where new paragraphs start. The number of subjects who mark a boundary as a paragraph boundary is then taken as the score for boundary strength (Rotondo, 1984; Swerts, 1997). The result of such a procedure is a representation of the layered hierarchical structure of a text: boundaries indicated as paragraph boundaries by many subjects are given a high position in the hierarchy; boundaries indicated as paragraph boundaries by few subjects are given a low position in the hierarchy. This procedure requires many subjects. Another procedure involves asking subjects to describe a well-structured object or task; the text structure is defined by reference to the object structure or task structure (Grosz, 1974; Terken, 1984; Swerts & Collier, 1992). These procedures all have in common that the intuitive knowledge of subjects concerning text structure is appealed to.

¹ An earlier version of this chapter was published in Den Ouden & Van Wijk (2000).

Some theoretical accounts of text structure are Story Grammar (Thorndyke, 1977), intention-based analyses (based on Grosz & Sidner, 1986), Rhetorical Structure Theory (Mann & Thompson, 1988), and PISA (Sanders & Van Wijk, 1996). The results of these accounts are fully connected trees representing both the hierarchical organized structure of a text and the labeled relations between the branches of the tree. Such theoretical accounts force analysts to reflect on their decisions and to make the reasons for these decisions explicit. If these procedures can be applied reliably, i.e., with high inter-subject reliability, the expertise of a single person is sufficient to obtain a hierarchical structure of a text.

Based on these two research traditions, four procedures were selected to examine the reliability of the analyses of the hierarchical structure of texts. In the intuition-based procedures, a group of naive annotators² indicated the paragraph boundaries in texts. Two variants were applied, a free task and a restricted task. In the free task, the number of boundary markers was free; in the restricted task, that number was fixed. The theory-based procedures used were Intention Based Analysis (so-called in this dissertation, henceforth IBA) and Rhetorical Structure Theory (henceforth RST), both well-known and widely used theories in discourse linguistics and computational linguistics, including computational approaches to prosody. The hierarchical structures resulting from IBA were labeled using so-called WHY?-labels, i.e., intentions, as formulated by the analysts. The hierarchical structures resulting from RST were labeled using the relation definitions, as formulated by the theory. The four procedures met the requirements of general applicability and numeric scoring of hierarchical levels as mentioned above. In this chapter a test of the reliability of the four procedures is described. The characteristics of the four procedures are presented in Table 2.1.

Table 2.1 Characteristics of four procedures for analyzing text structure

	INTUITION BASED		THEORY BASED	
PROCEDURE	free task	restricted task	IBA	RST
INSTRUCTION	indicate boundary markers (number is free)	indicate boundary markers (number is fixed)	specify on basis of WHY?-labels	specify on basis of explicit relation definitions
ANALYSIS	in groups		by individuals	
RESULT	unlabeled tree		labeled tree	

In the last ten years, the subjectivity of analyses has been a point of interest in computational linguistic and cognitive science studies of discourse and dialogue (Carletta, 1995; Carletta, Isard, Doherty-Sneddon, Isard, Kowtko, & Anderson, 1997; Condon & Cech, 1995; Flammia, 1998). For intuition-based procedures, reliability only holds for the derived structures that are obtained by adding up the annotations of individual subjects. With regard to the intuition-based

² In this study the term ‘annotator’ was reserved for persons applying intuition-based procedures, and the term ‘analyst’ was reserved for persons applying theory-based procedures. The general term used for both was ‘subjects’.

procedures, evidence that people have clear and reliable intuitions about discourse boundaries is provided by Bond and Hayes (1984), Hearst (1997) and Passoneau and Litman (1993, 1997). With regard to IBA, researchers have addressed the issue of its reliability by examining the agreement between annotators on categorical annotations of text-structural features such as the location of ‘segment beginnings’, ‘segment finals’, and ‘segment medials’ (Hirschberg & Grosz, 1992; Hirschberg & Nakatani, 1996) and the ‘beginnings of new intentions’ (Passoneau & Litman, 1993). Researchers assessing agreement on categorical labels ignore the hierarchical positions of segments in the whole text structure. For instance, there may be agreement on a categorical label like ‘segment beginning’, but the ‘beginnings’ may be embedded at different levels in the hierarchical structure. In these studies of the agreement on categorical labels, agreement between labelers on the multilayered hierarchical structures of texts was not assessed. With regard to RST, the requirement of reliability of these analyses was settled by a discussion between analysts leading to a consensus analysis of the text. Bateman and Rondhuis (1997:19), for instance, reported good inter-analysts reliability for RST, but they based that statement on a consensus reached after discussion between the analysts. A analysis based on consensus, however, does not show whether all analyses were represented equally or whether the analysis of the person with the highest status or the most forceful personality won. Not until recently did researchers address the question of inter-coder reliability of RST when analysts work independently (Den Ouden, Van Wijk, Terken, & Noordman, 1998; Marcu, Romera & Amorrtortu, 1999; Den Ouden & Van Wijk, 2000) or when texts are analyzed automatically (Marcu, 2000, Marcu & Echihiabi, 2002).

The research traditions of the four procedures are reasonably well established, and the reported experiences of reliability are generally good, but the procedures have not yet been compared directly with each other. The aim of the study described in this chapter was to determine whether analysts come up with the same hierarchical structures for the same texts when they apply a particular procedure, independently of each other.

2.2 Text structure analyses

In this section, each of the four procedures is explained and examples of results are presented.

2.2.1 Intuition-based procedures

When working with texts that were not specifically constructed and manipulated for the purpose of the research in studies of prosodic characteristics of text structure, researchers have predominantly applied intuitive, subjective procedures to annotate text structure (Rotondo, 1984; Swerts, 1997). The essence of these intuition-based procedures is that naive subjects indicate how a text is organized by making ‘annotations’ in the text. The annotations of individual subjects are markers (e.g., bars) that indicate major transitions in the text, for example, paragraph boundaries. Annotators may give the same weight to all paragraph boundaries, making a binary decision between ‘paragraph boundary’ (bar) or ‘no paragraph boundary’ (no bar); or they may give different weights to paragraph boundaries, for example, making a ternary distinction between ‘major boundary’ (double bar), ‘minor boundary’ (single bar), and ‘no boundary’ (no bar); or

they may make an n -ary scalar decision (by attributing scores). Furthermore, the annotators may be constrained in the number of boundary markers that they may annotate.

An essential feature of this procedure is that the hierarchical structure of the text is obtained by adding the number of individual subjects who mark each particular boundary as a paragraph boundary. Boundaries indicated as paragraph boundaries by many people are, therefore, considered strong boundaries: in a graphical hierarchical representation, such boundaries are located at high levels. Boundaries indicated as paragraph boundaries by few people are considered weak boundaries: in a graphical hierarchical representation, such boundaries are located at low levels. The hierarchical structure is unlabeled in that the subjects do not indicate the kind of relation that holds between text parts.

Two intuition-based procedures were applied in this study. In the free task, the annotators were free to decide on the number of boundary markers; in the restricted task, the number was fixed. An example of a result of the free task is presented in Table 2.2 and of the restricted task in Table 2.3. The restriction was to indicate four boundaries. The sample text on which these results are based is presented in Table 2.6. The texts presented did not contain paragraph makers and punctuation marks, except question marks. The original Dutch text is presented in Appendix A.

Table 2.2 Number of subjects ($n_{\max}=17$) in the free task who indicated boundaries as paragraph boundaries in the sample text

1-2	0	7-8	0	13-14	0	19-20	0	25-26	0	31-32	0
2-3	7	8-9	0	14-15	7	20-21	7	26-27	4	32-33	1
3-4	2	9-10	5	15-16	0	21-22	0	27-28	0	33-34	14
4-5	0	10-11	1	16-17	10	22-23	17	28-29	0	34-35	0
5-6	11	11-12	2	17-18	0	23-24	0	29-30	11	35-36	1
6-7	0	12-13	0	18-19	0	24-25	5	30-31	0	36-37	2

Table 2.3 Number of subjects ($n_{\max}=52$) in the restricted task who indicated boundaries as paragraph boundaries in the sample text

1-2	0	7-8	0	13-14	11	19-20	0	25-26	0	31-32	0
2-3	3	8-9	0	14-15	6	20-21	2	26-27	5	32-33	2
3-4	4	9-10	6	15-16	0	21-22	0	27-28	0	33-34	43
4-5	0	10-11	0	16-17	25	22-23	51	28-29	0	34-35	1
5-6	35	11-12	0	17-18	0	23-24	2	29-30	5	35-36	1
6-7	1	12-13	4	18-19	0	24-25	0	30-31	0	36-37	1

2.2.2 Theory-based procedures

Intention Based Analysis

According to Grosz and Sidner's theory of discourse structure (1986), three components of text structure must be distinguished: linguistic structure, attentional state, and intentional structure³. The linguistic structure consists of the sequence of the utterances. The attentional state consists of the dynamic record of the entities and attributes that are salient during a particular part of the text. In Grosz and Sidner's terms, this record is called 'stack', which expresses the focus of attention. The attentional state changes during the process in which the discourse unfolds, because speakers or writers may interrupt the main stream of the discourse (called 'push') or may take it up again (called 'pop'). Changes in linguistic structure and attentional state are dependent on the 'intentional structure' of the text; this structure consists of intentions or 'discourse segment purposes' (DSPs) underlying the discourse, and relations between DSPs. The basic idea is that speakers or writers have one or more particular intentions when they produce discourse. In order to express the segments as much as possible in accordance with these intentions, speakers or writers order and combine the segments with other segments in such a way that their purposes are communicated optimally. Hearers or readers, for their part, recognize the reason why a segment is produced, and they know that all segments are related to that purpose in some way and contribute to conveying that purpose. The purposes are organized hierarchically. Two kinds of relations account for the hierarchical structure in texts: dominance and satisfaction-precedence relations. In a dominance relation, a Discourse Purpose (DP) dominates one or more Discourse Segment Purposes (DSPs). In a satisfaction-precedence relation, a certain DSP2 can only be satisfied when a certain DSP1 has preceded it.

IBA is formulated as a procedure in the manual developed by Nakatani, Grosz, Ahn, and Hirschberg (1995). According to that manual, annotating text structure is equivalent to recognizing the speaker's underlying intentions. The analyst starts to identify the overall purpose of the text, which is comparable to formulating a title or headline. Purposes are described in terms of so-called WHY?-labels at the various levels of the text. The annotation of WHY?-labels is similar to making an outline of the discourse, although a WHY?-label captures not only the content of a (part of a) text, but also the speaker's or writer's reasons for letting the hearer or reader know that (part of) text. The dominance and satisfaction-precedence relations are visually expressed using indentations in the text: WHY?-labels occur at various hierarchical levels. The manual explicitly formulates some instructions for identifying these relations, but the hierarchical segmentation criterion in IBA, being based on the speaker's intentions, leaves room for personal interpretation.

Table 2.4 presents an IBA structure of the sample text presented in Table 2.6. The WHY?-labels indicate, 'What is the purpose of this section?' The hierarchical structure is built up using indentations. The figure shows that the analyst considered the text to consist of four main parts: 1-22; 23-33; 34-35; and 36-37. Segments 1 and 2 dominate the five sub-segments 3-5, 6-9, 10-13, 14-16, and 17-22. These sub-segments discuss different themes of Bolkestein, a well-known

³ Grosz & Sidner (1986) do not address the meaning of discourse, since "an adequate theory of discourse meaning needs to rest at least partially on an adequate theory of discourse structure" (p.198)

Dutch politician, and in that respect they are subsidiary to the purpose of segments 1 and 2. Although the sub-segments are not related to each other, each text part is directly related to the text part consisting of segments 1 and 2.

Table 2.4 IBA structure of sample text

level	Label on basis of WHY?-label	range
1	WHY? Bolkestein wants to keep pace with the US	1-2
2	WHY? Bolkestein talks on TV about the failure of policy in relation to minorities	3-5
2	WHY? Bolkestein has a strategy	6-9
2	WHY? Despite this strategy, no accusation can be made	10-13
2	WHY? He capitalizes on his activities to prove his point that normal immigrants don't need help	14-16
2	WHY? Bolkestein argues on TV for ending special grants for minorities	17-20
3	WHY? A conclusion: WHY? not abolish all government support?	21-22
1	WHY? Oprah Winfrey produced same results as Bolkestein	23
2	WHY? Winfrey wanted to show that support helps blacks who are deprived to improve	24-26
3	WHY? Plan failed	27-29
2	WHY? Other aid organisations are angry while others claim 'aid doesn't help'	30-33
1	WHY? Bolkestein's and Winfrey's actions lead to conclusion that support doesn't help	34-35
1	WHY? Real conclusion: support and motivation, not support or motivation	36-37

Rhetorical Structure Theory

Applying RST (Mann & Thompson, 1988) results in a multilayered hierarchical structure of a text, in which rhetorical relations between sentences are explicitly made, and nuclei and satellites are distinguished. RST is used to analyze texts as follows. First, the analyst splits the text into elementary units or segments on the basis of criteria specified by the theory. The segments are essentially clauses, where a clause is defined as (a part of) a sentence that contains a finite verb. Exceptions to this criterion are clausal subjects and complements, and restrictive relative clauses that are considered parts of their host clauses rather than separate segments. Second, the analyst groups the individual segments into text spans and specifies the rhetorical relations between adjacent text spans, in this way creating a hierarchical representation of the text. RST explicitly describes the rhetorical relations between segments. There are about 25 relations, for example, Evidence, Background, and Solutionhood. The relations are defined in terms of conditions on the nucleus, on the satellite, and on the combination of nucleus and satellite, and in terms of the effect on the reader. The nucleus is the central part of a text span and the satellite is peripheral. In Table 2.5 and Figure 2.1, the basic concepts of RST are illustrated using the Evidence relation. In example (1), segment 1 is the nucleus (represented by a vertical line); segments 2 and 3 are satellites (represented by an outgoing arrow).

Table 2.5 Definition of Evidence relation in terms of RST

EVIDENCE	
Constraints on N:	R might not believe N to a degree satisfactory to W
Constraints on S:	R believes S or finds it credible
Constraints on the N + S combination:	R's comprehending S increases R's belief of N
The effect:	R's belief of N is increased
Locus of the effect:	N

Note. N means nucleus; S means satellite; R means reader; W means writer

- (1)
1. The form was too difficult to fill in for this group of people.
 2. Almost everyone made mistakes in it
 3. and a large number of people did not even send it back.

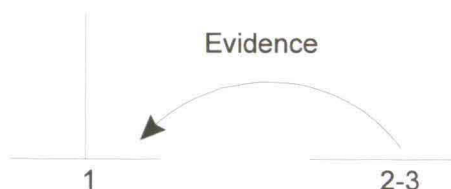


Figure 2.1 Label on basis of relation definition

The schema presented in Figure 2.1, with one nucleus and one satellite, is the most common one. Some schemas are multi-nuclear, like Contrast, Joint, and Sequence; in these relations, the text spans are equally important and there can be more than two text spans.

Analyzing texts using RST involves top-down and bottom-up analysis at the same time. In general, the process starts in a top-down way: the analyst divides the whole text into the largest text spans by finding the major boundary in the text, and determines what rhetorical relation exists between the text spans separated by that boundary. These text spans are in turn decomposed into smaller text spans and rhetorical relations between them are labeled, until finally the level of individual segments is reached. During the analysis, the arguments for one relation definition or another become explicit. The bottom-up process of analysis works the other way around: the analyst joins two individual segments to form a labeled rhetorical relation, thereby creating a text span; this text span is in turn joined with another segment or text span to form a rhetorical relation, and so on. Both strategies, top-down and bottom-up, are applied simultaneously until the whole text is analyzed successfully. The resulting structure is a labeled hierarchical representation of the text. Figure 2.2 presents an RST structure of the sample text (see Table 2.6; original Dutch text in Appendix A).

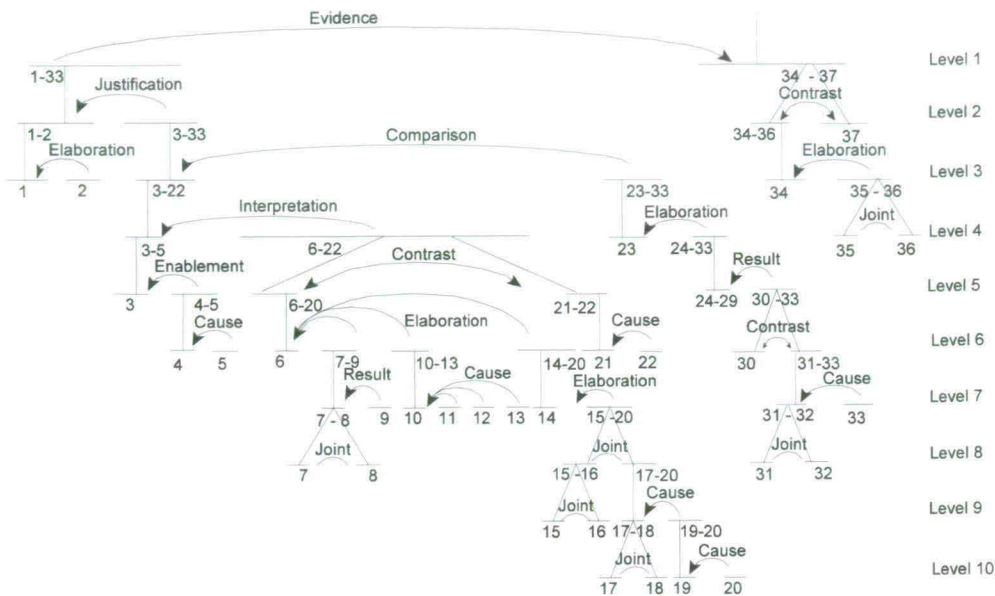


Figure 2.2 RST structure of sample text

The arrows in the figure connect those parts of the text between which some rhetorical relation holds. Each vertical line indicates a nucleus. The numbers under the horizontal lines indicate the segments that form a text span. Figure 2.2 shows that the text consists of 37 segments. The relation between text span 1-33 and text span 34-37 is characterized by Evidence. This means that the analyst considered 1-33 to be evidence for 34-37 based on the definition of the Evidence relation. Text span 34-37 is the statement to be believed and, therefore, it is the nucleus; text span 1-33 is the satellite that is intended to increase the reader's belief of the nucleus. One level lower in the hierarchy, text span 3-33 is a Justification of text span 1-2. One level lower, the segments are related to each other by way of an Elaboration: segment 2 elaborates the statement of the nucleus, segment 1, and so forth.

To compare the reliability of the four procedures, the hierarchical structures resulting from them were graphically represented in the same format such that the individual segments were at the bottom level. Figures 2.3a to 2.3d present the hierarchical structures of the sample text as they resulted from the four procedures. The figures show that the strongest boundary was between segments 22 and 23 following the intuitive procedures, whereas IBA located the three strongest boundaries between segments 22 and 23, between segments 33 and 34, and between segments 35 and 36, and RST located the strongest boundary between segments 33 and 34. The outcomes of the four procedures were clearly not the same.

Table 2.6 Sample text

segment	
1	the Netherlands keeps up with the United States of America once again
2	at least the Netherlands as imagined by Bolkestein, leader of the liberal party
3	last week he talked in the television programme Network about the failure of the policy in relation to minorities
4	he was allowed to appear in that programme
5	because he had published a booklet of interviews with successful Muslims
6	this is a well-trying Bolkestein strategy
7	curry favor with the Muslims
8	show how great they are in your eyes
9	and after that only talk about how the Netherlands should treat its minorities much more severely to realize real integration
10	and nobody can accuse him of anything
11	because after all he is the man who introduced the Moroccan Oussama Cherribi to the liberal party and the Lower House
12	after all he is the man who wrote a book together with the Algerian professor of Islam, Mohammed Arkoun
13	and now he is the man who has published a book about successful Muslims
14	and he capitalizes on all these books and actions to prove his point
15	normal immigrants succeed without special support from the government
16	they don't need that at all
17	so in the television programme Network he also argued for putting a stop to special government grants and attention for minorities
18	such grants must be given to all people who have a poor social position
19	because support doesn't really work
20	people who are really willing will succeed on their own
21	but Mr Bolkestein, why not abolish all government support at the same time?
22	money down the drain, isn't it?
23	in the United States, Oprah Winfrey by an opposite action unfortunately produced the same results as Bolkestein did here
24	in contrast, she just wanted to demonstrate that some extra support really helps black Americans who are socially deprived to lead a decent existence
25	she spent one million dollars
26	and set up an enormous organization full of educationalists, psychologists, and other experts, to help seven black families back on their feet
27	the plan failed miserably
28	Oprah stopped
29	when one of the women who had an enormous burden of debt refused to get rid of her mobile telephone
30	now the real aid organizations are furious; so much money for so few people
31	while the rest of America says 'See? you can't help these people'
32	and eagerly point to Oprah herself
33	look at her, she succeeded, and without any support, didn't she?
34	Bolkestein and Oprah, opposite actions with the same result and apparently the same conclusion
35	support doesn't work
36	motivation, that's what it's all about
37	when will some smart aleck hit upon the idea that maybe it is a matter of support and motivation rather than support or motivation?

Source: Column 'The state of the media' by Joan de Windt in radio program 'Tulpen en olijven', 23 May 1997

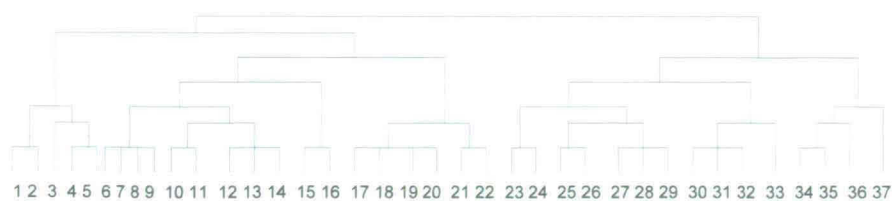


Figure 2.3a Hierarchical structure delivered using the free task

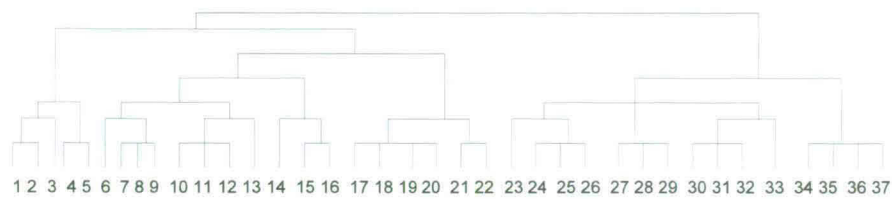


Figure 2.3b Hierarchical structure delivered using the restricted task

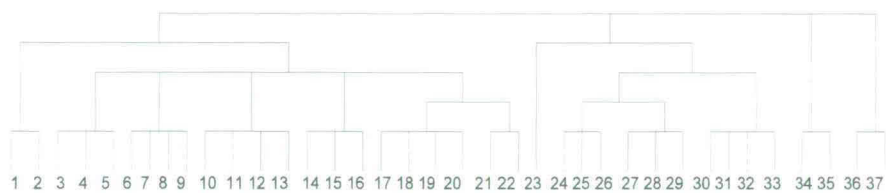


Figure 2.3c Hierarchical structure delivered by IBA

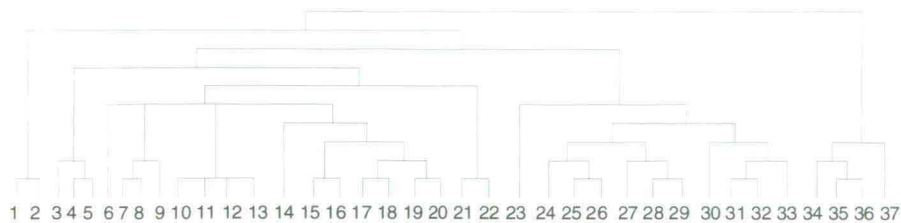


Figure 2.3d Hierarchical structure delivered by RST

2.3 Method

The methodology used to assess the reliability of the procedures is described in this section. Reliability was examined for the hierarchical levels of text structure only, not for the specification of the WHY?-labels or labels of the rhetorical relations.

2.3.1 Text material

Natural texts were selected instead of constructed or well-known texts from the literature. The text material consisted of four transcriptions of texts that were originally broadcast on Dutch radio. The texts were presented to the subjects for analysis in written form only, because the analysis of text structure had to be entirely independently of prosody: the spoken versions may contain cues that would cause the subjects to make particular choices when analyzing the texts (Grosjean, 1983; Gee & Grosjean, 1983; Hirschberg & Grosz, 1992; Hirschberg & Nakatani, 1996; Van Donzel & Koopmans, 1995; Swerts, 1997). In the studies reported in the following chapters, the text structures obtained in this study will be related to the prosodic realizations of these texts by the original speakers who were also the authors of the texts.

Two texts were news reports about actual events, one about Clinton's visit to Rome (Text I) and one about Berlusconi's problems (Text II). They were read aloud by two different news reporters from abroad by telephone. The other two texts were commentaries on actual events, one about the use of pocket telephones (Text III) and one about policy in relation to minorities both in the Netherlands and in the United States (Text IV, sample text). They were read aloud in the studio of the radio station by the authors of the texts. The news reports may be considered descriptive, narrative texts; they narrate a sequence of events relating to Clinton and Berlusconi. These texts were organized sequentially in that the events were reported successively. The commentaries may be considered argumentative texts; the aim is to give an opinion on a particular topic. These texts were organized more hierarchically than the descriptive texts in that they contained one central statement which was supported by the other text parts.

The texts were split into segments on the basis of the RST criteria. The length of segments ranged from five to thirty words. The descriptive texts were shorter than the argumentative texts. Text I contained 25 segments, Text II 28, Text III 35, and Text IV 37. These segmented texts were given to the subjects for analysis.

2.3.2 Procedures

2.3.2.1 Free task

Analysts. Seventeen naive subjects participated in this task. Their ages ranged from 17 to 59 years.

Task. The annotators were asked to indicate using bars the boundaries between segments which they considered to be important boundaries. They were free to decide how many bars to use, i.e., how many boundaries to indicate. The annotators were encouraged but not obliged to

differentiate between boundaries of different weight by using two bars for ‘strong boundary’ and one bar for ‘weak boundary’. Few annotators made a distinction between boundaries of different weights. For these annotators, only the ‘strong boundaries’ were analyzed.

2.3.2.2 Restricted task

Analysts. Fifty-two students following the ‘Text and Communication’ programme at the Faculty of Arts at Tilburg University participated in this task. Their average age was about 21 years.

Task. The students were asked to indicate using bars the boundaries between segments which they considered to be important boundaries. They were limited with respect to the number of boundaries they could annotate per text. For Texts I and II, both being shorter than Texts III and IV, the participants had to annotate three important boundaries; for Texts III and IV, four boundaries were to be marked.

2.3.2.3 Intention Based Analysis

Analysts. Three persons analyzed the texts using IBA. They were all senior researchers and well experienced in IBA. Two of them were affiliated (at that time) with AT&T, Florham Park, New York; one was affiliated with Boston University. Experts rather than novice users of IBA were preferred, because high-quality analyses were required. There were two options: getting expert users from abroad or training people in the Netherlands. The first option was chosen, because the training of people was considered too laborious. The four Dutch texts were translated into English by a professional translator.

Task. The analysts were asked to analyze the texts carefully using the practical manual by Nakatani, Grosz, Ahn, and Hirschberg (1995). They were asked to create a hierarchical structure of each text using indentations, and to analyze WHY?-labels indicating the discourse segment purposes. They did not communicate with each other about the task. No limitations were imposed on the amount of time taken to complete the task.

2.3.2.4 Rhetorical Structure Theory

Analysts. Six persons analyzed the texts using RST. They were text linguists who were well experienced in RST: two Ph.D. students and four senior researchers. They were members (at that time) of the Discourse Studies Group of Tilburg University. RST experts were chosen instead of novice users, because high-quality analyses were required. Training of novices was considered too laborious.

Task. The analysts were asked to analyze the texts carefully according to Mann and Thompson (1988). They had to create a complete hierarchical structure of each text with nuclei and satellites, and to give labels to the rhetorical relations between the segments. The analysts did not

communicate with each other about this task. The time they could spend on the task was unlimited.

2.3.3 Data analysis

Scoring of hierarchical levels

In each procedure, scores were given to the boundaries in the graphical hierarchical representations of the text structures in the following way: for each boundary, count the number of branching nodes dominating the segments separated by that boundary until a common dominating node is reached.

In the RST structure represented in Figure 2.3d, for example, the boundaries between segments 1 and 2, 4 and 5, 7 and 8, 10 and 11, and so forth, are all scored as 1, because they are immediately dominated by a common node. The boundary between segments 29 and 30 is scored as 5, because there are five branching nodes dominating the segments separated by that boundary before a common dominating node is reached: three nodes dominate segment 29, one node dominates segment 30, and one common node dominates the boundary between segments 29 and 30. The boundary between segments 33 and 34 is scored as 9, because there are nine branching nodes dominating the segments separated by the boundary between segments 33 and 34: six nodes dominate segment 33, two nodes dominate segment 34, and one common node dominates the boundary between segments 33 and 34. The scores express the weights which were given to the boundaries such that higher values are associated with 'more important'.

Statistical analysis

Agreement between subjects was computed in two ways. First, weighted kappa statistics for evaluating agreement concerning categorical judgements⁴ was used (Cohen, 1960; Cohen, 1968; Carletta, 1995; Siegel & Castellan, 1988; Popping, 1996). Second, Spearman's rank correlations were used to examine the pairwise relations between individual RST analysts and between individual IBA analysts.

2.4 Results

Table 2.7 presents the values of kappa measures of agreement⁵ for the procedures per text. Kappas are evaluated on the basis of their values: kappas between .61 and .80 signify substantial agreement, between .41 and .60 moderate agreement, and between .21 and .40 fair agreement (Rietveld & Van Hout, 1993: 221).

⁴ An objection may be raised against the use of kappa since the scores of the boundary levels in the theoretical approaches and the decisions to make major boundaries in the intuitive approach were not obtained independently of each other. However, the kappa statistic is generally accepted as a standard measure for assessing annotation reliability (Carletta, Isard, Isard, Kowtko, Newlands, Doherty-Sneddon, & Anderson, 1995; Flammia & Zue, 1995; Shimojima, Katagiri, Koiso & Swerts, 1999; Van Herwijnen & Terken, 2001).

⁵ Thanks to Roel Popping (University of Groningen, Department of Sociology) for his computations of kappa using the AGREE program, and for his helpful comments.

Table 2.7 Kappa measures of agreement for each text

		Intuition-based procedures		Theory based-procedures	
		free task (k=17)	restricted task (k=52)	IBA (k=3)	RST (k=6)
Text I	(n=24)	0.59	0.55	0.35	0.56
Text II	(n=27)	0.52	0.52	0.53	0.52
Text III	(n=34)	0.37	0.49	0.15	0.52
Text IV	(n=36)	0.46	0.55	0.41	0.68

A value of .40 was taken as the lower boundary; the intuition-based procedures reached this standard for three of the four texts. In the free task, the kappas were lower for Texts III and IV than for Texts I and II. The annotators performing the free task had more difficulty in reaching agreement on the structures of the argumentative texts than on the structures of the descriptive texts (Texts III and IV versus Texts I and II). The free task was performed using lower agreement than the restricted task for two texts.

For two texts, IBA did not reach the standard of .40. The kappas of IBA were lower than those of the intuition-based procedures, except for text II. For all texts, the kappas of RST reached the standard of .40. The kappas of RST were higher than those of IBA, and in the same range or higher than those of the intuition-based procedures.

To examine the agreement in more detail, pairwise Spearman's rank correlations were computed for the IBA and RST structures found by the three and six analysts, respectively. Table 2.8 presents the range of pairwise Spearman's rank correlations between the six RST analysts and the three IBA analysts. For the intuition-based procedures, no correlations between hierarchical structures could be computed, because the hierarchical structure of a procedure was derived from the combined annotations of the individual subjects.

Table 2.8 Range of pairwise Spearman's rank correlations between the RST and IBA analysts

	IBA (k=3)	RST (k=6)
Text I	.29 - .68	.51 - .88
Text II	.49 - .82	.60 - .82
Text III	.09 - .54	.59 - .88
Text IV	.44 - .68	.76 - .95
Texts taken together	.43 - .50	.69 - .87

For RST, the mean of the pairwise correlations was .76 following Fisher's z conversion. Each correlation was significant. For IBA, the mean of the pairwise correlations was .46 following Fisher's z conversion. One third of the correlations did not reach significance.

2.5 Conclusion and discussion

Kappas were computed as measures of agreement between the subjects of each of the four procedures. Pairwise Spearman's rank correlations were computed as measures of the relations between the analysts of IBA and between the analysts of RST. Based on the results of this study, a procedure for analyzing text structure has to be selected for the research on the relation between hierarchical structure and prosody, described in the following chapters.

Intuition-based procedures are commonly used in prosodic research on text structure. Figures 2.3a and 2.3b show that these procedures result in multilayered structures that are useful for prosodic research on texts. The restricted variant of the intuition-based procedure was applied with higher agreement than the free variant. Therefore, in using these procedures for the annotation of a particular text, it is better to restrict the number of paragraph boundaries subjects can indicate. Hierarchical structures delivered in this way can be based on moderate agreement.

Of the theory-based procedures, IBA did not perform as well as RST. This is a remarkable result, since the hierarchical structures resulting from IBA (see Figure 2.3c) are not as complex and elaborate as the structures resulting from RST (see Figure 2.3d). Given that IBA is less explicit than RST in defining the conditions under which relations between segments hold, and in that respect has a greater resemblance to intuition-based procedures, it was expected that the kappas for IBA would be relatively as high as the kappas for the intuition-based procedures, but this was not the case.

The agreement in RST was as good as the agreement in the restricted intuition-based procedure, and for one text even clearly better. RST gives, in addition to a reliable hierarchical structure, a huge amount of information about the ways in which the text parts are related to each other. This kind of information is completely lacking in the intuition-based procedure. The written texts presented to the subjects were not complete texts as they did not contain prosodic characteristics, but they did not contain written cues either which could signal the text's structural organization, such as capital letters at the beginnings of segments, punctuation marks at the endings of segments, indentations, and blank lines. Agreement between RST analysts would even have been better if they could have analyzed the texts with all cues available.

RST is used in the studies described in the following chapters as its reliability was found to be sufficiently high, even higher than that of the intuition-based procedures, and because it provided the most detailed analysis of the structure of the texts in terms of both hierarchy and rhetorical relations.

The better reliability of RST may be accounted for in two ways. First, in RST, the relation definitions are explicitly described in terms of conditions on the segments concerned and the relations between them (see Table 2.5). This is a major difference with IBA, because in IBA the WHY?-labels of the relations are mainly based on text summarizations. IBA does not prescribe a fixed set of relation names.

Second, partly as a result of the explicit definitions in RST, the analysis of a text in terms of RST is a laborious task. The mean time required by the six RST analysts was 139 minutes for Text I, 128 minutes for Text II, 123 minutes for Text III, and 105 minutes for Text IV. The difference with the other procedures is striking: the time required by the annotators of the

intuition-based procedures was twenty minutes for all four texts, and by the three IBA analysts the mean time required was 23 minutes for Text I, 17 minutes for Text II, 15 minutes for Text III, and 15 minutes for Text IV. This means that RST took over seven times more time than IBA. The time required to analyze a text indicates that an RST analyst processes a text in-depth: the highly specific relation definitions force an RST analyst to think about the text more thoroughly than an IBA analyst and annotators of the intuitive procedures do. The high agreement between the RST analysts shows that the great amount of time used was not wasted.

Using the text material and the hierarchical structures resulting from RST, the relation between text structure and prosody will be explored in the study described in Chapter 4. For the four texts, the level scores of the RST analyses were related to prosodic characteristics of the original speakers of the texts.

3

Reliability of pitch-range measurements

3.1 Introduction¹

Variation in pitch range is a conspicuous aspect of natural speech. It reflects the enthusiasm or emotional state of the speaker (Mozziconacci, 1998) and the position of an utterance in the text structure (Sluijter & Terken, 1993; Möhler & Mayer, 2001; Portes, Rami, Auran, & Di Christo, 2002; Den Ouden, Noordman & Terken, 2002, 2003). For research on variation in pitch range, a manageable and reliable method for measuring the pitch range is required. The reliability of two methods for measuring the pitch range of an intonation phrase² is investigated in this study. Figure 3.1 illustrates the basic concepts underlying these methods. It shows an imaginary pitch contour consisting of a sequence of pitch rises and falls. Pitch peaks coincide with the end of a pitch rise and the beginning of a pitch fall, and are associated with accented syllables. The terms *pitch* and *F0* are used exchangeably in this chapter.

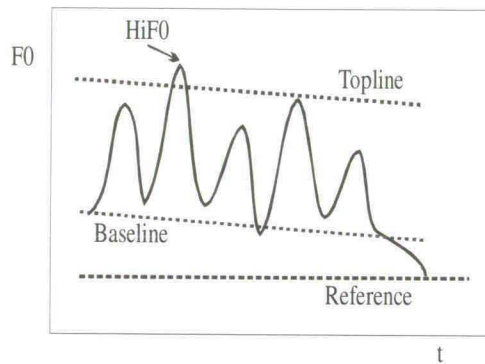


Figure 3.1 Imaginary, stylized pitch contour

The literature offers at least two approaches for characterizing the pitch range of an intonation phrase. One approach, proposed by Liberman and Pierrehumbert (1984) and also included in the influential ToBI framework (Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert, & Hirschberg, 1992), defines pitch range in terms of the distance between the *F0*-maximum of the phrase (*High F0*) and a minimum, the so-called *Reference*, which is considered to be constant for each speaker. In read-aloud speech, the *Reference* is usually reached at the end of an utterance (Liberman & Pierrehumbert, 1984); in spontaneous speech, it is reached at the end of topical units (Wichman, 1991). The *High F0* is the highest peak of the intonation phrase. In the approach of Liberman and Pierrehumbert, it is assumed that the *High F0* is under the control of the speaker and that the values of the other peaks in the intonation phrase are derived

¹ An earlier version of this chapter will be published in Den Ouden, Terken, Van Wijk & Noordman (accepted).

² In read-aloud text, each sentence coincides with an intonation phrase, except for sentences that consist of multiple clauses, where each clause is realized as a separate intonation phrase.

using a simple rule. That is, the value of the highest peak is a measure of the pitch range of the whole intonation phrase (Pierrehumbert, 1980, 1981; Liberman & Pierrehumbert, 1984).

A different class of models is connected with the IPO approach to intonation ('t Hart, Collier, & Cohen, 1990; see also Sorensen & Cooper, 1980 and Ladd, 1990). It is assumed that, in an intonation phrase, the peaks and valleys in the course of the pitch range may be captured by two gradually declining lines (*Topline*, *Baseline*). These lines are called declination lines. Similar to the model of Liberman and Pierrehumbert, the end frequency of the baseline is supposed to be more or less constant for a given speaker. In earlier versions of the IPO model, the topline and baseline were assumed to run parallel. Pitch range was then expressed as the distance between those lines. According to later versions of the model, speakers are able to vary the topline and baseline independently, and both the topline and the baseline are communicatively relevant (Ladd, 1990, 1993; Terken, 1993; Gussenhoven, Repp, Rietveld, Rump, & Terken, 1997). The difference with Liberman and Pierrehumbert (1984) is that, according to the IPO approach, variation of the baseline bears independent communicative relevance. The perceptual relevance of the baseline has been demonstrated in various experiments, for example, on the perception of prominence (Gussenhoven et al., 1997; Gussenhoven & Rietveld, 2000), on the perception of emotions (Mozziconacci, 1998), and on the identification of an utterance as a question or a declarative sentence (Haan, Van Heuven, Pacilly, & Van Bezooijen, 1997). The perceptual relevance of the baseline suggests that the characterization of the variation in pitch range is not complete when the variation of the baseline is left out of consideration.

3.2 Pitch-range measurements

The theoretical frameworks differ with respect to the prosodic parameters that need to be measured; either the F0-maximum or the course of declination. It should be noted that a simple metric for pitch range is considerably more complicated in the IPO approach than in the approach proposed by Liberman and Pierrehumbert, owing to the independent variation of the topline and the baseline. Even the measurement of the F0-maximum is problematic, since it has to be characterized as much as possible in terms of its perceptual relevance, and not only in terms of its physical characteristics.

The F0-maximum is the highest F0-value in the accented syllable of an utterance. The syllable that has physically the highest F0, however, is not necessarily the syllable that is perceived as the syllable with the highest pitch. For example, the second peak in Figure 3.1 is physically the highest peak, but it is not necessarily associated with an accented syllable; instead, the fourth peak may be the accented syllable (Pierrehumbert, 1979). In general, linguistic knowledge is needed to determine whether syllables are accented or not. Also, listeners compensate for declination of the contour, so that a syllable with lower F0 later in the utterance may in fact be perceived to have higher pitch range than a syllable with higher F0 earlier in the utterance (Silverman, 1987). Automatic measurements are not (yet) capable of determining the F0-maximum perfectly since they lack the kind of linguistic knowledge needed, and even human listeners may disagree on it.

Another problem in measuring F0-maxima are microprosodic phenomena which affect the course of the pitch contour, such as specific consonants and vowels, and a creaky voice. For

example, under equal circumstances, the peaks of the vowels [i] and [e] are higher than the peaks of [a] and [o]. Based on their linguistic knowledge, trained phoneticians are able to correct for these influences while measuring F0-maxima, but automatic methods are not able to do this. This may lead to inaccurate results for both the values and the locations of the peaks in the utterance. Some researchers have attempted to correct for microprosodic influences. Thorsen (1979), for example, developed a system that accounts for the identities of the vowels.

Characterization of the declination lines of an intonation contour is yet more complicated. Toplines do not have to connect all peaks of the contour, but only those peaks which are associated with *accented* syllables. Moreover, the peaks and valleys are often not aligned neatly on two straight lines (see Figure 3.1). Therefore, in order to determine topline and baseline, deviations have to be weighted. Both for human judges and for automatic procedures, it is difficult to find out which peaks and valleys have to be taken into account to fit the declination lines, and how the deviations from the ideal line have to be weighted. A well-known procedure for measuring the lines automatically is by fitting linear regression lines. Lieberman, Katz, Jongman, Zimmerman and Miller (1985) characterized a pitch contour using a single line. They based the linear regression line of an utterance on all points of the contour. Haan et al. (1997) characterized a pitch contour using two linear regression lines representing the topline and the baseline. These lines were fitted in two steps: first, the central linear regression line was computed in the same way as was done by Lieberman et al. (1985); then, separate regression lines were computed for the points above this central line and for the points below it. The perceptual relevance of the variation in the pitch contour was not taken into consideration in computing these lines.

Both human judges and automatic procedures require linguistic knowledge to measure pitch range. This knowledge is beyond the reach of current automatic systems. Human judges have to be relied on, because of their perceptual capabilities and their (implicit) linguistic knowledge. Nevertheless, the question needs to be answered whether human listeners characterize a pitch contour reliably, since their linguistic knowledge and their perceptions and judgements may be different. In this study, human judges are asked to estimate the slopes of declination lines and the values and locations of F0-maxima. With respect to the peak-based approach, the relevant parameters are the value of the F0-maximum relative and its location in the utterance; for the lines-based approach, the relevant parameters are the beginning and ending of both topline and baseline, the slope of the lines, and the relation between the slopes. We expect that the reliability between judges might be affected by the length of an utterance (Swerts, Strangert, & Heldner, 1996; Cooper & Sorensen, 1981), for instance, because longer utterances contain more information than shorter ones. This factor is taken into consideration. Finally, the human judgements are compared with automatic measurements (see also Portes & Di Christo, 2003).

3.3 Method

3.3.1 Judges

Five trained phoneticians participated in the experiment: four senior researchers and one junior researcher. Four of them were affiliated (at that time) with the research programme for Spoken Language Interfaces of the Institute for Perception Research (IPO) of Technische Universiteit Eindhoven; one was affiliated (at that time) with the Phonetic Laboratory of Leiden University. They were all familiar with the speech-processing program in which the speech material was presented to them.

3.3.2 Speech material

Forty utterances were selected from the four read-aloud texts of which written transcripts were used in the studies reported in Chapter 2. The texts had been broadcast originally on Dutch radio. Two texts were telephone news broadcasts and two were commentaries, each read aloud by a different speaker. There were three male speakers and one female speaker. The speakers were experienced in reading aloud for the radio. The shortest text contained twenty-five utterances³, the longest 37. From each text, ten utterances were selected which on screen showed clear declination patterns: an utterance consisting of only one intonation phrase and not containing an end rising. The utterances were split up into two groups: half of the utterances were longer than 5 seconds (mean: 6.11 sec.; sd: 1.03 sec.), and half were shorter than 5 seconds (mean: 3.82 sec.; sd: 0.64 sec.). The forty utterances were presented to the judges in four lists. Each list consisted of ten utterances taken from a single speaker. The order of the four lists was the same for all judges. At the start, four practice utterances were added to familiarize the judges with the speakers' voices.

3.3.3 Procedure

The speech material was presented to the judges in the speech-processing program Gipos (Graphical Interactive Processing Of Speech, developed at IPO, Eindhoven University; <http://www.ipoe.tue.nl/ipoe/gipos>). The judges performed the task on their own computers⁴. Figure 3.2 presents an example of a pitch contour as the judges saw it on their screens. The speech-processing program enabled them to listen to the utterances as often as they needed to, and to examine the amplitude envelope, and so forth. The task was self-paced.

³ The term 'utterance' is loosely used in this chapter. In a strict sense, 'clause' is meant. Within the framework of text analysis as applied in the following chapters, the clause is referred to using the term 'segment'.

⁴ Our thanks to Leo Vogten for programming the task (Department of Electrical Engineering, Technische Universiteit Eindhoven, The Netherlands)

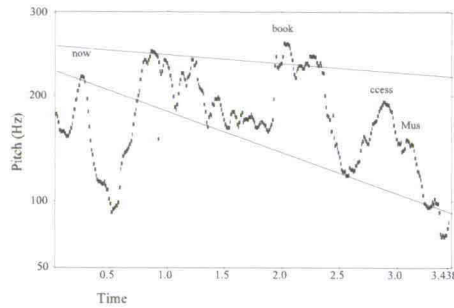


Figure 3.2 The pitch contour of the utterance:
*"....and now he is the man who has published a
 book about successful Muslims"*

Three tasks were performed per utterance, independently of each other, on the basis of visual and auditory inspection. First, the judges had to fit a baseline; the line was shown on the screen and could be modified until the final decision was made. Second, after the screen was cleared, the judges had to fit a topline of the utterance in the same way. Third, after the screen was cleared again, the judges determined the value and location of the F0-maximum, defined as the highest F0-value associated with a pitch accent. The judges were instructed to ignore consonantal perturbations and breaking voices. In this way, six parameters per utterance were obtained for each judge: the values of the beginning and ending of the topline (in hertz), the values of the beginning and ending of the baseline (in hertz), the value of the F0-maximum (in hertz), and the location of the F0-maximum (in seconds). The temporal locations of beginning and ending were automatically determined using the amplitude envelope.

3.4 Results

3.4.1 Effect of length on pitch-range measurements

As mentioned in the introductory section, longer utterances contain more information that is relevant to fitting declination lines than short ones, so that an effect of length might be expected for the declination lines. Therefore, the relation between the pitch-range parameters and length was examined first. If the parameters differed with regard to length, length would have to be taken as a separate factor in the analyses of agreement. Table 3.1 presents the values of the pitch-range measurements averaged over the five judges separately for long and short utterances. Using the parameter values of topline and baseline indicated by the judges, the slopes of the declination line were computed for each utterance, i.e., the difference between beginning and ending was divided by the duration of the utterance. For example, a negative slope of 9.2 indicate that the declination was 9.2 hertz per second. The relation between the slopes was also computed, i.e., the slope of the topline was divided by the slope of the baseline. A value of 1 meant, therefore,

that the lines ran parallel; a value higher than 1 meant that the lines converged, i.e., the topline declined more than the baseline; a value lower than 1 meant that the lines diverged, i.e., the baseline declined more than the topline.

Table 3.1 Pitch-range measurements for long and short utterances (standard deviations between brackets)

			Short (n=20)	Long (n=20)
Value of F0-maximum	(in hertz)		231 (49.3)	241 (50.3)
Location of F0-maximum	(in seconds)		1.62 (0.9)	1.94 (1.0)
Topline declination	beginning	(in hertz)	225 (41.4)	237 (43.6)
	ending	(in hertz)	169 (28.4)	179 (34.0)
Baseline declination	beginning	(in hertz)	136 (20.3)	134 (19.9)
	ending	(in hertz)	102 (14.4)	101 (11.2)
Slope	topline	(in Hz/sec)	-15.4 (8.6)	-9.2 (3.6)
	baseline	(in Hz/sec)	-9.6 (4.2)	-5.5 (2.2)
Relation between slopes	(slope _{topline} /slope _{baseline})		1.79 (1.5)	1.90 (0.9)

Two-way ANOVAs were performed with Judge as within-group factor and Length as between-group factor, and each of the pitch-range parameters as the dependent variables. Length did not affect any of the parameters (all F 's < 1), except the slope of the topline ($F(1,38)=7.66$, $p<01$, $\eta^2=.17$) and the slope of the baseline ($F(1,38)=12.58$, $p<.001$, $\eta^2=.25$). For both topline and baseline, the values of the beginnings and endings were independent of the length of the utterance (in accordance with 't Hart et al., 1990). As a logical consequence of this constancy, the slopes of the declination lines were steeper in short utterances than in long ones. Of all interactions, only the interactions between Length and Judge were significant for the location of the F0-maximum ($F(4, 152) = 3.00$, $p<.025$, $\eta^2=.07$) and for the slope of the topline ($F(4,152)=2.45$, $p<.05$, $\eta^2=.06$). The judges differed with regard to the location of the F0-maximum and the declination of the topline for short and long utterances. Although the factor Length proved to be of little influence, this factor was included in the subsequent analyses of reliability in order to assure maximum control over its effect.

3.4.2 Agreement between human pitch-range measurements

Agreement between the judges who measured pitch-range parameters was examined both in a relative and in an absolute way. Relative agreement was investigated by correlating the scores of the five judges. Absolute agreement was investigated by testing differences between the five judges.

3.4.2.1 Relative agreement

For each pitch-range parameter, agreement was determined using Cronbach's alpha and the pairwise correlations between pairs of judges. By convention, values of Cronbach's α above .70 are called *adequate* and above .80 *good*. A value of .90 among five judges may be called *excellent*. The data are presented in Table 3.2.

Table 3.2 Agreement between judges concerning pitch-range measurements and the range of the correlations

		Cronbach's α	Range of correlations
Value of F0-maximum		.98	.86 to .98
Location of F0-maximum		.86	.23 to .93
Topline declination	beginning	.96	.79 to .93
	ending	.92	.61 to .91
Baseline declination	beginning	.95	.78 to .88
	ending	.95	.75 to .89
Slope	topline	.90	.37 to .89
	baseline	.91	.58 to .80
Relation between slopes (slope _{topline} /slope _{baseline})		.34	-.40 to .63

Cronbach's α was above .90 in almost all cases. The judges were a homogeneous panel. It should be noted, however, that the value of Cronbach's α is highly dependent on the number of judgments (Van Wijk, 2000: 220). As the number of judgments was high (40 items in this case), the result may be flattered. Therefore, the pairwise correlations between judges were examined as well. The correlations were high (at least .86) for the values of the F0-maximum. The correlations were somewhat lower (with .61 as minimum) for the values of beginnings and endings of the lines. The ten pairwise correlations for the F0-maximum were compared with the ten pairwise correlations for the beginnings of the topline and the baseline using Wilcoxon Signed Ranks tests. The pairwise correlations of the measurements of the F0-maximum were higher than the pairwise correlations of the measurements of the other pitch-range parameters (topline begin: $z=2.46$, $p<.025$; baseline begin: $z=2.81$, $p<.005$). The value of the F0-maximum was measured more reliably than the values of the beginnings of both lines. The correlations for the location of the F0-maximum (.23 as minimum) and for the slopes (.37 as minimum) were low. There was no agreement about the relation between the slopes. From the Cronbach's α 's, it may be concluded that the judges agreed strongly in relative terms (Rietveld & Van Hout, 1993: 203). As described in the following section, it was examined whether they agreed absolutely as well.

3.4.2.2 Absolute agreement

For each pitch-range parameter, the differences between the judges were tested for significance using an analysis of variance with Judge as within-group factor (five levels: coded A, B, C, D, E) and Length of the utterance as between-group factor (two levels: short, long). In several analyses a second within-group factor was included. Table 3.3 presents for each judge the mean of the F0-maxima, the locations of the F0-maximum in relation to the beginnings of the utterance, the beginnings and endings of topline and baselines, the slopes, and the relations between the slopes, computed over the forty utterances.

Table 3.3 Pitch-range measurements per individual judge

			Judge				
			A	B	C	D	E
F0-maximum		(in hertz)	236	239	236	240	228
Location F0-maximum		(in seconds)	1.44	1.50	2.25	1.60	2.47
Topline declination	beginning	(in hertz)	233	236	234	235	218
	ending	(in hertz)	197	171	166	182	156
Baseline declination	beginning	(in hertz)	146	136	137	122	133
	ending	(in hertz)	110	101	95	95	106
Slope	topline	(in Hz/sec)	-7.9	-14.7	-14.9	-11.6	-13.5
	baseline	(in Hz/sec)	-7.9	-7.9	-9.2	-5.9	-5.9
Relation between slopes (slope _{topline} /slope _{baseline})			1.24	2.29	1.74	1.46	2.48

There was an effect of Judge on the F0-maximum ($F(4,152)=4.16$, $p<.005$, $\eta^2=.10$). A posteriori comparisons showed that of ten pairwise comparisons, only two were significant: judge E differed from judges B and D.

There was an effect of Judge on the location of the F0-maximum ($F(4,152)=13.86$, $p<.001$, $\eta^2=.27$). Judges A, B, and D placed the F0-maximum at approximately 1.5 seconds, judges C and E did so at approximately 2.5 seconds. There was little agreement about the exact location of the F0-maximum. For the 40 utterances, the five judges decided 17 times (43%) unanimously. In 11 cases, there was one judge who disagreed with the other four; the difference between the two indicated locations was 1.83 seconds on average. In 12 cases, only three judges agreed; the difference between the two indicated locations of the F0-maxima was 2.70 seconds on average. When three of the five judges identified the same peak as the F0-maximum, the location was at 0.80 seconds on average. When four of the five judges considered the same peak to be the F0-maximum, the location was at 2.01 seconds on average; when all judges indicated the same peak, the location was at 1.91 on average. There was more disagreement on F0-maxima which were near to the beginning of the utterance than on F0-maxima which occurred later in the utterance.

For the evaluation of the values of the beginnings and endings of the lines, Location was added as within-group factor (four levels: beginning and ending of topline, beginning and ending of baseline). There was an interaction between Location and Judge ($F(12,456)=18.74$, $p<.001$,

$\eta^2=.33$). The way the judges disagreed differed for the four locations. One-way analyses of variance for each location separately showed that the judges differed significantly with regard to their estimations of each of the four locations (Topline beginning: $F(4,152)=6.39$, $p<.001$, $\eta^2=.12$; Topline ending: $F(4,152)=24.56$, $p<.001$, $\eta^2=.39$; Baseline beginning: $F(4,152)=31.56$, $p<.001$, $\eta^2=.45$; Baseline ending: $F(4,152)=44.64$, $p<.001$, $\eta^2=.54$). For the beginning of the topline, Judge E differed from the four other judges: four out of ten pairwise comparisons were significant. For the ending of the topline, all pairwise comparisons were significant, except between judges A and D, and between judges B and C: eight of the ten pairwise comparisons were significant. For the beginning of the baseline, all judges differed pairwise, except judges B, C, and E: seven of the ten pairwise comparisons were significant. For the ending of the baseline, all judges differed pairwise, except judges C and D, and A and E: eight of the ten pairwise comparisons were significant.

Two remarks can be made based on the pairwise comparisons of the five pitch-range measurements. First, there was no atypical judge, because the judges who differed were not the same persons each time. Second, the least number of differences in pairwise comparisons between judges was found for the F0-maximum.

For the slopes of the lines, Line was added as within-group factor (two levels: topline, baseline). For the slopes, there was an interaction between Line and Judge ($F(4,152)=14.86$, $p<.001$, $\eta^2=.28$). The way the judges disagreed differed for the topline and the baseline. One-way analyses of variance for each line separately showed an effect of Judge for both the topline ($F(4,152)=14.22$, $p<.001$, $\eta^2=.27$) and the baseline ($F(4,152)=12.77$, $p<.001$, $\eta^2=.25$). For the topline, judge A differed from all others, and judges B and C differed from D. For the baseline, judges A, B, and C differed from judges D and E. Again the differences were not caused by an 'atypical' judge.

For the relation between the slopes, there was a small effect of Judge ($F(4,152)=2.18$, $p=.07$, $\eta^2=.05$). Judge A differed from judges B and E, and C differed also from E. All judges had a score higher than 1, i.e., all of them indicated that the declination of the topline was steeper than the declination of the baseline. The courses of both lines had a range from slightly converging (1.24) to strongly converging (2.48). For two judges, the lines did not differ significantly from a parallel course (A: $t(39)=1.04$, $p=.31$; D: $t(39)=0.67$, $p=.50$). The scores of the other three judges deviated significantly from 1; their toplines had a steeper declination than their baselines (B: $t(39)=3.37$, $p<.005$; C: $t(39)=5.03$, $p<.001$; E: $t(39)=11.49$, $p<.001$).

The problem of fitting declination lines is illustrated by the following example. As can be seen from Figure 3.2, the peaks and valleys cannot always be captured by clear trend lines. As the data show, the judges came to different measurements. Figures 3.3a, 3.3b, and 3.3c illustrate the possible extent of these differences. The figures show the declination lines indicated by three judges for the utterance, 'Nederland loopt weer eens gelijk op met de Verenigde Staten van Amerika' ('The Netherlands keeps up with the United States of America once again'). One of the judges determined that the declination lines had a parallel course (Figure 3.3a), another judge determined that the lines converged (Figure 3.3b), and the third judge determined that the lines diverged (Figure 3.3c). The consistency between the judges in the fitting of declination lines may have been greater if more constraints had been imposed on the judges, but the validity of such

constraints with respect to measuring pitch range would require further research, which was beyond the scope of this thesis.

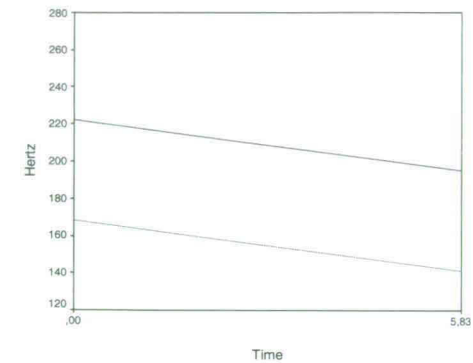


Figure 3.3a Parallel declination lines
($b_{top} = -7.65$, $b_{base} = -7.65$, $b_{top}/b_{base} = 1.00$)

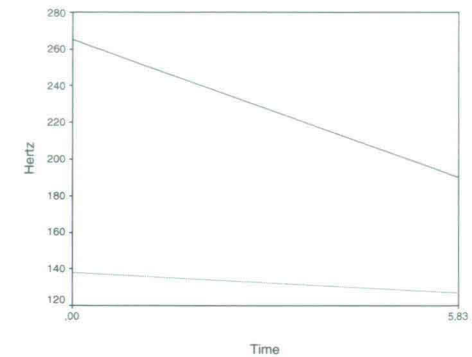


Figure 3.3b Converging declination lines
($b_{top} = -21.25$, $b_{base} = -3.12$, $b_{top}/b_{base} = 6.82$)

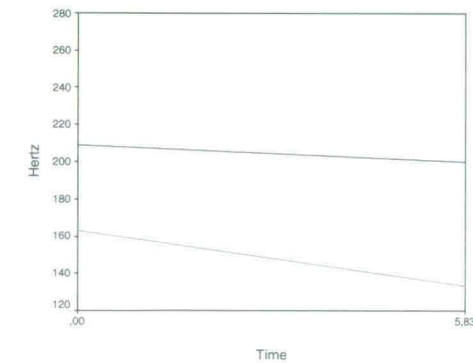


Figure 3.3c Diverging declination lines
($b_{top} = -2.55$, $b_{base} = -8.50$, $b_{top}/b_{base} = 0.30$)

3.4.3 Agreement between human judges and automatic measurements

Relative agreement between the judges was high for all prosodic parameters, except for the relation between the slopes. The analyses of absolute agreement showed that there were few pairwise comparisons where the judges differed regarding the value of the F0-maximum, whereas there were many more pairwise differences for the other prosodic parameters. Therefore, to examine the connection with automatic measurements, the F0-maximum seemed to be the best candidate. Table 3.4 presents the correlations between the automatic measurements and the scores of the F0-maximum determined by the five judges.

Table 3.4 Correlations between the scores of the judges and the automatic measurements of the F0-maximum (value in hertz; location related to beginning of utterance in seconds)

	Judge					Majority
	A	B	C	D	E	
Value of F0-maximum	.97	.93	.93	.99	.90	.99
Location of F0-maximum	.78	.84	.49	.91	.25	.91

The correlations between the automatic measurements and the scores of the judges were high for the F0-maximum. The values of three of the five judges correlated almost perfectly with the value which was measured automatically. Strikingly, in 38 of the 40 utterances, the automatic measurements were slightly higher than the human measurements (6.5 hertz on average). The human judges possibly compensated for microprosodic influences in accented syllables based on their linguistic knowledge.

The correlations were much lower for the location of the F0-maximum on the time-axis, and they varied strongly per judge from .25 to .91. There was a correlation of .91 between the score of the majority, i.e., the location of the F0-maximum which was determined by at least three of the five judges, and the value which was measured automatically. The automatic measurement differed only slightly from the majority judgement in 37 of the 40 utterances (less than 80 milliseconds). In the other three cases, the human judges indicated a completely different location of the F0-maximum: twice earlier in the utterance, once later.

3.4.4 Relevance of F0-maximum as characterization of pitch range

Of the two methods of measuring pitch range, using the pitch peak in the contour (F0-maximum) or using the beginnings and endings of the declination lines, the F0-maximum seems to be the more reliable. The question is then whether we can be sure that all relevant information is captured. Even though the judges reached a high degree of agreement in identifying the F0-maximum, it does not automatically follow that the F0-maximum captured all the information that is relevant to pitch-range variation.

In order to find out whether the values of the beginnings and endings of the declination lines might code information that is not captured by the F0-maximum, we explored the correlations between these pitch-range parameters. In order to compute correlations between the parameters, average values were computed across the five judges for all utterances. The slope of the topline did not correlate with its ending ($r=.02$, $p=.89$), but it did with its beginning ($r=-.41$, $p<.01$). The same applied for the slope of the baseline (ending: $r=-.12$, $p=.46$; beginning: $r=-.57$, $p<.001$). Inspection of the data suggests that this was due to the end frequencies of the topline and baseline being more or less constant. Hence, the variation of a declination line is approximated best by the value at the beginning of the line. These onset values turned out to correlate strongly with the values of the F0-maxima (topline: $r=.90$, $p<.001$; baseline: $r=-.77$, $p<.001$). In other words, a substantial part of the variation of both declination lines may be accounted for by the variation of the F0-maximum. This suggests that the F0-maximum captures all or most of the information about the pitch range of an utterance, and that the other parameters can be computed by rule, as

was argued by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984). It should be noted, though, that this reasoning applies only in the context of read-aloud descriptive texts. The variation of the baselines of spontaneous, more expressive speech utterances has been found to code independent information, as variation of the baseline with a constant topline gives rise to the perception of different emotions (see, for example, Mozziconacci, 1998).

3.5 Conclusion and discussion

Pitch range may be characterized either in terms of the pitch peak in the contour (F0-maximum) or in terms of the beginnings and endings of the declination lines. The reliability of these pitch-range parameters varied strongly. In terms of relative agreement the judges agreed strongly on all prosodic parameters, except the relation between the slopes: Cronbach's α was above .90 for each of the pitch-range measurements. The pairwise correlations showed that the F0-maxima were determined more consistently than the beginnings and endings of the declination lines. Overall, the relative agreement between the judges was strong for all parameters.

The absolute agreement turned out to be weak for all parameters, although less weak for the F0-maximum. Not all, but most judges did agree on the values of the F0-maximum. They did not agree at all on the beginnings and endings of the declination lines, nor on the slopes of the declination lines, and the location of the F0-maximum. The number of pairwise differences between judges was smaller for the F0-maximum than for the other parameters. There was also good agreement between the values of the F0-maxima indicated by the judges and those found using the automatic procedure; the correlations were above .90.

The high degree of relative agreement and the lower degree of absolute agreement suggest that individual judges may have a bias for measuring pitch-range parameters, but that this bias is systematic. Therefore, for measuring pitch range in a relative way, for instance, when the pitch range of all utterances in a text has to be measured, use of the measurements of one judge would suffice.

The F0-maximum met the requirements of a manageable and reliable criterion for measuring pitch range well. The judges agreed strongly about its value, especially in relative terms. It can be measured automatically in a sound way, and the score is independent of the length of an utterance. Although the declination lines give additional information for spontaneous, more expressive speech, the F0-maximum is a useful measure of pitch range in read-aloud speech.

4

Prosody of hierarchy: An exploration

4.1 Introduction

The central topic of this dissertation is the reflection of text structure in prosody. In earlier research on prosody in texts, only two levels in text structure were distinguished: sentences and paragraphs. The focus was on differences in the prosodic marking of boundaries between sentences and boundaries between paragraphs. It was shown, for example, that first sentences of paragraphs have a longer preceding pause and higher pitch range than sentences within paragraphs, and that parentheticals and final sentences of paragraphs are articulated with lower pitch range and at a faster rate than sentences at other locations in the text (Brubaker, 1972; Lehiste, 1975; Silverman, 1987). In later research, more levels in text structure were distinguished, for example, based on judgments of boundary strength (Swerts, 1995, 1997) or categories of segment types in the text structure (Hirschberg & Grosz, 1992; Hirschberg & Nakatani, 1996; Van Donzel, 1999), or on theories of text analysis such as PISA (Schilperoord, 1996), Story Grammar (Noordman, Dassen, Terken, & Swerts, 1999) and Rhetorical Structure Theory (Noordman et al, 1999). In general, these studies showed that pause duration and pitch range are related to text structure in such a way that the durations of pauses and the heights of fundamental frequency gradually decrease as text-structural categories or hierarchical levels become more subtle. Prosodic realizations of hierarchical levels in text analyses as such have been investigated in Swerts (1997) and Noordman et al. (1999). In Swerts (1997), the annotation of hierarchical structure was based on the intuitions of naive language users, whereas a theoretically based analysis of text structure was thought to provide a more adequate and reliable account. In Noordman et al. (1999), the hierarchical structures of the texts were theoretically explicated, but the texts used were constructed texts which had been analyzed repeatedly in the literature. In this dissertation, both natural texts and theoretically based hierarchical structures are used to investigate the relation between text structure and prosody.

For text structure, reliable text analyses are used, based on the results of the study reported in Chapter 2. The RST analyses of the natural texts provide the multilayered hierarchical structures required for research on text prosody. A question not addressed in the study reported in Chapter 2 was the scoring of the levels of the text structures.

For prosody a reliable measurement of pitch range is used, based on the results of the study reported in Chapter 3. The F0-maximum was used as a relevant and reliable measure of the pitch range per segment. Pause durations preceding segments, F0-maxima, and articulation rate of segments are the prosodic features investigated in relation to the hierarchical levels.

For the relation between hierarchy in text structure and prosody, this study is explorative in two aspects. First, various ways to score the hierarchical structure of texts are investigated and the consequences are examined. We distinguish three procedures: top-down, bottom-up, and symmetrical. Second, various ways to relate hierarchical structure to prosodic realizations are explored. We distinguish two ways to approach the relation between hierarchy and prosody: absolutely and relatively. The hierarchical levels of a whole text structure may be scored absolutely on an interval scale and then related to their prosodic realizations, or the hierarchical levels of a pair of segments may be scored relatively on an ordinal scale, i.e., in terms of 'higher' and 'lower' and then related to their prosodic realizations. The hypothesis is that the higher the

levels are in the hierarchy, the more strongly they are realized prosodically. For the three prosodic parameters, we expect that the pauses preceding segments will be longer, the F0-maxima of the segments will be higher, and the articulation rate of the segments will be slower, the higher the segments in the hierarchical structure of the text.

Another factor taken into account in this study is the syntactic status of segments. The RST criteria for segmentation do not take into account the syntactic status of the segments. Segments can be simple main clauses, but also coordinate main clauses and subordinate clauses. The hierarchical structure of texts may be associated with the syntactic status of the segments (Cooper & Sorensen, 1977, Sorensen & Cooper, 1980). For example, main clauses may occur more frequently at high levels and subordinate clauses more frequently at low levels of a text structure, and they may be realized prosodically in different ways. To avoid any confounding of text structure with syntactic status of the segments, syntactic status is controlled for.

4.2 Method

4.2.1 Text material

The four texts introduced in the studies described in Chapter 2 were used again. The four segmented texts were presented to six RST experts. Each person analyzed the four texts in terms of RST. For the study reported in this chapter, one analysis of each text would have been sufficient, but the six resulting text analyses were not exactly identical. Therefore, the level scores of the boundaries of the four texts were based on the average level scores of the six text analyses of each text. In the statistical analyses, these average scores were rounded off to the nearest integer. They were the independent variable.

4.2.2 Three procedures for scoring hierarchical level

There are several ways to score the levels of an hierarchical structure. Here, we distinguished a top-down, bottom-up, and symmetrical procedure of scoring. Each scoring procedure is illustrated on the basis of one of the six RST analyses of a sample text. The analysis selected was chosen, **because it gave the highest mean pairwise correlation between the six analysts.** Table 4.1 presents the sample text (the original Dutch text is presented in Appendix B); Figure 4.1 presents the text structure.

In the top-down procedure, the levels of the boundaries between segments were scored from top to bottom in the RST analysis. The procedure was as follows. Score 1 was attributed to the levels of the highest boundaries in the hierarchy, i.e., the boundaries between the largest text spans of the texts; score 2 to the levels of the boundaries one level lower in the hierarchy; and so forth. The scores of the levels are shown in the right margin of Figure 4.1. For example, boundary 4-5 is scored as 1, boundary 6-7 as 2, boundary 5-6 as 3, boundary 7-8 as 4, and so forth. In the top-down scoring, high boundaries were given low scores and low boundaries high scores. A consequence of the top-down procedure was that there were few boundaries at level 1. The scoring depended largely on the length of the text span the segments were part of: two boundaries

at the bottom of the RST analysis may have very different scores. For example, boundary 2-3 was scored as 3 and boundary 12-13 was scored as 9, though they were equal in that they were both at the lowest levels of the hierarchy. A consequence was that there were not many high scores. Therefore, results of statistical analyses are based on a small number of observations for the very low and very high boundaries.

In the bottom-up procedure, the levels of the boundaries between segments were scored from bottom to top. Figure 4.2 presents the bottom-up representation of the RST analysis presented in Figure 4.1. In this representation, the segments are aligned at the bottom of the analysis.

The bottom-up procedure was as follows. Score 1 was assigned to the lowest boundaries in the hierarchy, i.e., the boundaries at the bottom of the representation; score 2 to the boundaries one level higher in the hierarchy; and so forth. The scores of the levels of the boundaries are shown in the right margin of Figure 4.2. For example, the boundary between segments 2 and 3, and the boundary between segments 3 and 4 were scored as 1; the boundary between segments 1 and 2 as 2; the boundary between segments 10 and 11 as 3; and the boundaries between segments 4 and 5, and between segments 19 and 20 as 9. In the bottom-up representations of the hierarchical structures, low boundaries were expressed by low scores and high boundaries were expressed by high scores. A consequence of the bottom-up procedure was that, compared with the top-down procedure, the number of boundaries at the lowest level was extremely high. A problematic consequence of the bottom-up procedure was a risk that it would lead to ambiguous scores because of the asymmetry of the branches through which the scoring proceeded. For example, the boundary between segments 13 and 14 could be scored both from the left branch (containing segments 9 to 13) and from the right branch (containing segments 14 to 16). Scoring from the left would lead to a score of 5, scoring from the right to a score of 3. In these cases, the highest score of both was taken, for example, the boundary between segments 13 and 14 was scored as 5.

The symmetrical procedure was already introduced in the study reported in Chapter 2. The scores of the levels of the boundaries were based on the number of nodes which dominated a boundary, including the dominating node itself. The bottom-up representations of the RST analyses were used for this procedure.

The procedure was as follows: for each boundary, determine the superordinate node connecting the two segments adjacent to the boundary, count the number of subordinated nodes including the connecting node itself; the number of nodes is the score of the boundary. In Figure 4.2 it can be seen, for example, that the boundary between segments 19 and 20 was scored as 7, since the superordinate node of segments 19 and 20 dominated six nodes, five at the left side and one at the right side; including the connecting node itself, this brought the score to 7. The boundary between segments 2 and 3 was scored as 1; the boundary between segments 1 and 2 as 2; the boundary between segments 6 and 7 as 4; the boundary between segments 4 and 5 as 5, and so forth. The symmetrical procedure overcame the drawback of the top-down procedure in that the boundaries between the individual segments at the bottom of the analysis all had the same scores. It also overcame the drawback of the bottom-up procedure in that the scores did not depend on the branch along which the scoring proceeded. A drawback of the symmetrical

procedure was that high boundaries could be scored differently although they were at the same height of the structure. For example, the boundary between segments 4 and 5 was scored as 5 and the boundary between segments 19 and 20 as 7, whereas these boundaries had the same scores in the top-down (both scored as 1) and bottom-up procedures (both scored as 9). Another oddity was that, in the symmetrical procedure some high boundaries had lower scores than boundaries lower in the hierarchy. For example, the boundary between segments 6 and 7 was scored as 4 and the boundary between segments 13 and 14 was scored as 6, whereas the boundary between segments 6 and 7 was closer to the top node. No measures were taken in this study to overcome this odd consequence of the symmetrical procedure.

Table 4.1 Sample text

segment	
1	this morning, Clinton started his first day in Rome as if he was at home, with a bit of jogging
2	beside him panted the American ambassador in Rome
3	who while running told him about the sights of the eternal city
4	and a little bit behind the security officers ran with guns in their shorts
5	after changing clothes, Clinton paid his respects to President Scalfaro
6	with whom he talked about democracy and human rights
7	after that the conversation with the pope was not so easy
8	it was especially about the UN conference in Cairo in September about population growth and development
9	in the prepatory document for that conference the UN opted for contraceptives and abortion as means to reduce the population growth in the Third World
10	Clinton agrees with that document
11	and it provoked the pope
12	for months now the pope has been conducting a crusade against this approach to the population problem
13	which in his view amounts to murder and the destruction of the family
14	President Clinton assured the pope that for Americans the family also occupies a central position
15	and he made it clear that he takes a document of the American Catholic church very seriously
16	in that document it says that Catholics in the United States will never agree with the opinion of their president about the population problem
17	in practice Clinton and the pope hardly agreed
18	at a press conference Clinton admitted as much
19	but he added that there was agreement about the necessity of sustainable development in the Third World
20	the press conference was held after a conversation between Clinton and prime minister Berlusconi
21	the Italian prime minister said that there was absolutely no question of fascism in his government it's a pseudo-problem, he said
23	an opinion poll shows that only nought point four percent of the Italians are nostalgic about fascism
24	moreover, he said, all my ministers are democrats
25	and they all believe totalitarianism must be fought

Source: *Radiojournal* by Jan van der Putten, 4 April 1994

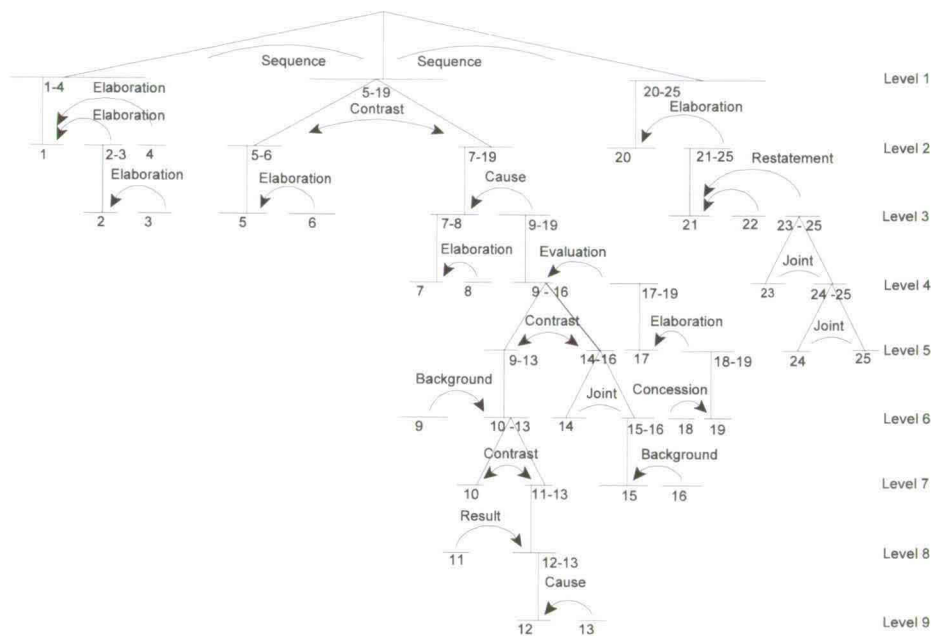


Figure 4.1 RST analysis of sample text

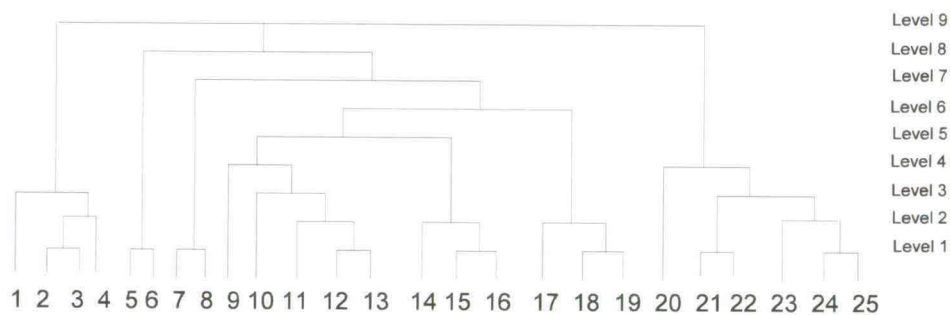


Figure 4.2 Bottom-up representation of RST analysis in Figure 4.1

Since each procedure seems to have its own bias, it was examined to what extent they differed or simply overlapped. Table 4.2 presents the scores of all boundaries of the sample text resulting from the three procedures. In the top-down procedure, score 1 was the highest boundary; in the bottom-up and symmetrical procedure, score 1 was lowest boundary. Note that the top-down and bottom-up procedures were not exactly opposites, i.e., the lowest scores in the top-down procedure did not completely correspond to the highest scores in the bottom-up procedure. For example, in the top-down procedure, the boundary between segments 6 and 7 and the boundary between segments 20 and 21 were both scored as 2, whereas in the bottom-up procedure the boundary between segments 6 and 7 was scored as 8 and the boundary between segments 20 and 21 as 4. Although the symmetrical and bottom-up procedures had much in common, especially at lower levels, there was a considerable difference between the two procedures at higher levels. For example, the boundary between segments 4 and 5 was scored as 9 in the bottom-up procedure and as 5 in the symmetrical procedure; and the boundary between segments 6 and 7 was scored as 8 in the bottom-up procedure and as 4 in the symmetrical procedure.

Table 4.2 Scores of the boundaries of the sample text in the three procedures

	top-down	bottom-up	symmetrical		top-down	bottom-up	symmetrical
1-2	2	3	3	13-14	5	5	6
2-3	3	1	1	14-15	6	2	2
3-4	2	2	2	15-16	7	1	1
4-5	1	9	5	16-17	4	6	5
5-6	3	1	1	17-18	5	2	2
6-7	2	8	4	18-19	6	1	1
7-8	4	1	1	19-20	1	9	7
8-9	3	7	5	20-21	2	4	3
9-10	6	4	2	21-22	3	1	1
10-11	7	3	2	22-23	3	3	3
11-12	8	2	2	23-24	4	2	2
12-13	9	1	1	24-25	5	1	1

To examine to what extent the three procedures were related, correlations were computed. For the four texts taken together, the correlation between the top-down and bottom-up procedure was $-.55$; between the top-down and symmetrical procedure, $-.44$; and between the bottom-up and symmetrical procedure, $.84$ (for each correlation, $p < .01$). Since the procedures did not deliver the same scores for hierarchy, the relation with prosody was examined for the three scoring procedures separately.

The scoring of the hierarchical levels of texts remained a problem. The three procedures reflected the hierarchical structure of texts in some way, but each had its shortcomings in representing all quantitative aspects of the hierarchical structure. In the literature, the notions of text structure and levels effect are often used in a non-problematic and self-evident way, but once

one starts to quantify the levels of a text analysis, it becomes obvious that there are various ways to do it and that each way has its own disadvantages and benefits. Because the scoring procedure itself was considered a substantial part of the problem of the relation between text structure and prosody in this study, all the three scoring procedures were explored in relation to prosody.

4.2.3 Speech material

The four selected texts were originally broadcast on Dutch radio. They were read aloud by four speakers, three male and one female, all used to speaking in public. The speakers were the authors of the texts. It was assumed that as authors they were most aware of the structure of their texts. The texts were segmented using the RST criteria and analyzed in terms of RST. The prosodic features of the segments of the read-aloud texts were measured. For pause duration, the beginnings and endings of the absences of speech between segments were indicated by hand. After that, the speech programma computed the exact durations in milliseconds. Pitch range, operationalized as the F0-maximum, was measured for each segment automatically in hertz (Den Ouden & Terken, 2001). The pitch contour of each individual segment was inspected for pitch measurement errors using an analysis-resynthesis procedure¹. Pitch-measurement errors consisted mostly of voiced-unvoiced errors and incorrect outliers in the speech signal. These errors were corrected in the speech signal by hand before the automatic measurements procedure was applied. The articulation rate of segments was measured in number of phonemes per second. Articulation rate based on the number of phonemes per second differs from that based on the number of syllables per second. When the production of speech, and not the perception, is considered as in the studies reported in this dissertation, the number of phonemes per second has been found to be a reliable measurement for the tempo of speech that is produced at a normal rate (Caspers, 1994). Time of articulation differs from time of speaking in that pauses within sentences are excluded, whereas they are included in time of speaking. When read-aloud speech, and not spontaneous speech, is considered as in the studies reported here, articulation rate is a better measurement since including the pauses within segments would misrepresent the tempo of speech of a read-aloud segment. In the studies reported in this dissertation the durations of pauses within segments were measured, and subtracted from the total duration of segments.

Table 4.3 presents the original prosodic measurements of the segments of the four texts. Pause durations of 0 milliseconds occurred in the speech material, meaning that some sequences of RST-defined segments were read aloud without intervening pauses.

¹ Thanks to Leo Vogten for programming automatic means to measure F0-maxima (Department of Electrical Engineering, Technische Universiteit Eindhoven, The Netherlands)

Table 4.3 Characteristics of the original prosodic measurements for the four texts together

		minimum	maximum	mean	standard deviation
preceding pause	(in milliseconds)	0	1290	540	240
F0-maximum	(in hertz)	156	378	238	48
articulation rate	(in phonemes per second)	12.1	19.8	15.3	1.6

Table 4.4 presents the original prosodic measurements of the segments per text, i.e., speaker.

Table 4.4 Characteristics of the original prosodic measurements for each text, i.e., speaker (standard deviations in brackets)

		preceding pause (in milliseconds)	F0-maximum (in hertz)	articulation rate (in phonemes per second)
Text I	(male)	570 (320)	218 (25)	15.1 (1.3)
Text II	(male)	600 (240)	219 (21)	16.1 (1.3)
Text III	(male)	440 (180)	208 (25)	15.8 (1.7)
Text IV	(female)	550 (210)	294 (45)	14.4 (1.5)

The considerable differences between the individual speakers were eliminated using standard scores instead of the original prosodic measurements. Per speaker, the original measurements were transformed into standard scores. In the following sections, only standard scores of the prosodic measurements are reported.

4.2.4 Two approaches for relating text structure and prosody

Two approaches were selected to explore the relation between the scores for hierarchical level and prosodic realizations: an absolute approach and a relative approach.

In the absolute approach, the levels of the hierarchical structures, defined on an interval scale, were directly related to the mean prosodic realizations of those levels. The relation between level scores and prosodic realizations was explored using the three procedures for quantifying the hierarchical levels. In the relative approach, the levels of the hierarchical structures were considered ordinal scores: the levels of pairwise-related segment boundaries were scored in terms of 'higher' and 'lower', and as such related to the mean prosodic realizations of the higher and lower boundaries. Two variants of the relative approach were distinguished, the 'procedure of linear adjacency' and the 'procedure of hierarchical adjacency'. In the procedure of linear adjacency, the relation between level scores and prosodic realizations was explored using the three procedures for quantifying the hierarchical levels. In the procedure of hierarchical adjacency, the relation between level scores and prosodic realizations was explored using the top-down procedure for quantifying the hierarchical levels.

An overview of the characteristics of the absolute and relative approaches is presented in Table 4.5. The two variants of the relative approach are explained below Table 4.5.

Table 4.5 Characteristics of absolute and relative approaches of relating hierarchy and prosody

	ABSOLUTE	RELATIVE	
PROCEDURE		hierarchical adjacency	linear adjacency
SCORING	top-down bottom-up symmetrical	top-down	top-down bottom-up symmetrical
OPERATIONALIZATION	interval scores	ordinal scores	
ANALYSIS	whole text structure	pairwise-related segments	

In the procedure of linear adjacency, pairs of adjacent boundaries in the text were compared with each other with regard to their prosodic features. For example, the prosodic features of the boundary between segments 1 and 2 were compared with the prosodic features of the boundary between segments 2 and 3; the prosodic features of the boundary between segments 2 and 3 were compared with the prosodic features of the boundary between segments 3 and 4; the prosodic features of the boundary between segments 3 and 4 were compared with the prosodic features of the boundary between segments 4 and 5, and so forth. For each pair of adjacent boundaries, it was determined which of the two adjacent boundaries was higher and which was lower in the hierarchical structure. The boundary scores used were the average boundary scores of the six analyses per text. Adjacent boundaries with equal level scores were not included in the analyses. Table 4.1 illustrates the procedure to determine the relative height of the scores of the adjacent boundaries. For example, the boundary between segments 1 and 2 had a higher score than the boundary between segments 2 and 3; and the boundary between segments 4 and 5 had a higher score than the boundary between segments 5 and 6, in all three procedures (in the top-down procedure, lower scores represented higher boundary levels). Figure 4.3 presents the two possible patterns of three consecutive segments and the relative height of the boundaries between them: the high-low pattern as depicted in Figure 4.3a, and the low-high pattern as depicted in Figure 4.3b. From the pair of adjacent boundaries between segments 1 and 2, on the one hand, and between segments 2 and 3, on the other hand, the boundary between segments 1 and 2 is the higher boundary in Figure 4.3a, whereas it is the lower boundary in Figure 4.3b.

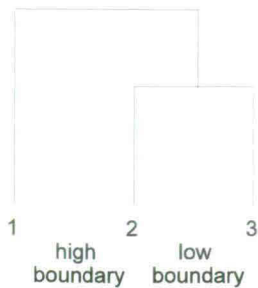


Figure 4.3a High-low pattern

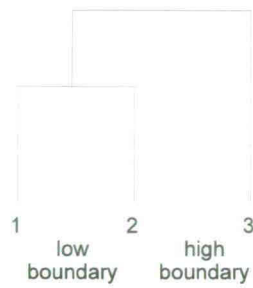


Figure 4.3b Low-high pattern

Table 4.2 shows that the high-low patterns of adjacent boundaries could differ for the three scoring procedures: for example, the score of the boundary between segments 9 and 10 was higher than the score of the boundary between segments 10 and 11 in the top-down and the bottom-up procedures, but the scores were equal in the symmetrical procedure. Because in the procedure of linear adjacency the average boundary scores were used, it was not possible to derive the relative height of adjacent boundaries directly from the graphical representations of the text analyses. To examine to what extent the three scoring procedures differed, the high-low patterns of the adjacent boundaries were determined for each of the three procedures and then pairwise compared. Four possible patterns could occur. They are illustrated on the basis of Table 4.2, although the scores in this table were derived from one particular text analysis and not from the average scores of the boundaries of the six analyses.

First, the high-low pattern could be the same for both scoring procedures: the same boundary would be the higher one and the same boundary the lower one in both scoring procedures. An example in Table 4.2 is the patterns of the boundary between segments 1 and 2 and the boundary between segments 2 and 3 for the bottom-up and symmetrical procedures. Second, the high-low pattern of the levels of the two adjacent boundaries could be the opposite for the scoring procedures: a particular boundary would be the higher one of a pair in one procedure, and the lower one in the other procedure. An example of this pattern is not found in Table 4.2. Third, in one of the two procedures, the pair of adjacent boundaries would not be involved because the boundaries had equal levels, whereas the boundaries had unequal levels in the other procedure. An example in Table 4.2 is the patterns of the boundary between segments 10 and 11 and the boundary between segments 11 and 12 for the bottom-up and symmetrical procedures: the pair is not involved in the symmetrical procedure because these adjacent boundaries have equal levels, whereas the pair is involved in the bottom-up procedure. Fourth, the pair of adjacent boundaries would not be involved in either procedure because the boundaries had equal levels in both procedures. An example of this pattern is not found in Table 4.2. Table 4.6 presents the frequencies of the four possible high-low patterns for the three procedures of linear adjacency, pairwise compared.

Table 4.6 Frequencies of high-low patterns for the three procedures of linear adjacency, pairwise compared

	top-down versus bottom-up	top-down versus symmetrical	bottom-up versus symmetrical
same high-low pattern	89 (76%)	77 (66%)	97 (83%)
opposite high-low pattern	0	2	1
one pair not involved in the analysis	23	35	13
neither pair involved in the analysis	5	3	6

Particularly because of the high frequencies in the third row of Table 4.6, which indicate that a considerable part of the pairwise comparisons did not overlap in the three procedures, the prosodic realization of the three procedures of quantifying text structure was explored. The three observations of opposite high-low patterns could be explained by the use of average scores per boundary instead of scores delivered by one particular text analysis.

In the procedure of hierarchical adjacency, pairs of boundaries within a particular branch of the hierarchical structure which have a direct relation of superordination and subordination were compared with each other with regard their prosodic features. For example, Figure 4.1 shows that the boundary between segments 4 and 5 dominates the boundary between segments 6 and 7, because the boundary between segments 4 and 5 is the boundary between text span 1 to 4, on the one hand, and text span 5 to 19, on the other hand, and because the boundary between segments 6 and 7 is, one level lower, a further division of the text span consisting of segments 5 to 19. In the same way, the boundary between segments 6 and 7 dominates the boundary between segments 5 and 6 and the one between segments 8 and 9; and the boundary between segments 8 and 9 dominated the boundary between segments 7 and 8 and the one between segments 16 and 17, and so forth. In the procedure of hierarchical adjacency, the determination of the higher and lower boundaries was based on the graphical representation of one RST analysis per text. It was not possible to rely on the average boundary scores, because in average scores of the boundary levels the particular relations of superordination and subordination were no longer defined. Therefore, from the six analyses which were available for each text, the one was selected that had the highest mean correlation with the other five text analyses (based on the symmetrical procedure of scoring). The four resulting text analyses, one for each text, originated from three different analysts.

4.3 Results

4.3.1 Effect of syntactic status on prosody

To check the effect of the syntactic status of the segments on the prosodic realizations, three syntactic classes of segments were distinguished: independent main clauses, main clauses which were the second part of a complex sentence consisting of two coordinate main clauses (called

coordinate main clauses), and subordinate clauses. Table 4.7 presents the mean standard scores of the prosodic parameters for each syntactic class.

Table 4.7 Prosodic characteristics in relation to three syntactic classes of the segments

		preceding pause	F0-maximum	articulation rate
independent main clauses	(n = 88)	0.36 *	0.16	0.04
coordinate main clauses	(n = 24)	-0.82	-0.33	-0.29
subordinate clauses	(n = 13)	-0.81	-0.50	0.27

Note: Based on 84 cases since a pause preceding the first segment of each text could not be measured

ANOVAs were run with syntactic class as independent variable and each of the three prosodic features as dependent variables. The effect of syntactic class was significant for pause duration ($F(2, 118) = 25.72, p < .001, \eta^2 = .30$) and the F0-maximum ($F(2, 122) = 4.41, p < .05, \eta^2 = .07$), but not for articulation rate ($F(2, 122) = 1.58, p = .21$). Pairwise comparisons in post-hoc analyses (Tukey's HSD procedure) showed that the pauses preceding independent main clauses were longer than pauses preceding clauses of the two other syntactic classes. The F0-maximum of the independent main clauses was higher than that of the subordinate clauses, but not higher than that of the coordinate main clauses.

On the basis of the results for preceding pauses and F0-maxima, independent main clauses were considered 'independent segments' and the other two classes 'dependent segments'. Table 4.8 presents for each syntactic class the mean standard scores of the preceding pause duration, F0-maximum, and articulation rate.

Table 4.8 Prosodic characteristics in relation to the syntactic classes of the segments

		preceding pause	F0-maximum	articulation rate
independent segments	(n = 88)	0.36*	0.16	0.04
dependent segments	(n = 37)	-0.82	-0.39	-0.09

Note: Based on 84 cases since a pause preceding the first segment of each text was not measured.

Pauses preceding independent segments were longer than pauses preceding dependent segments ($F(1, 119) = 51.87, p < .001, \eta^2 = .30$). The F0-maximum of the independent segments was higher than that of the dependent segments ($F(1, 123) = 8.61, p < .01, \eta^2 = .07$). Syntactic class did not affect articulation rate ($F < 1$).

The distribution of syntactic classes over hierarchical levels was examined, because an unequal distribution of syntactic classes over hierarchical levels may affect their effect on prosody. Table 4.9 presents for each level the distribution of the two syntactic classes, scored symmetrically.

Table 4.9 Distribution of syntactic classes over hierarchical levels scored symmetrically

		independent segments	dependent segments
level 4 and higher	(highest in the hierarchical structure)	24 (96%)	1 (4%)
level 3		18 (95%)	1 (5%)
level 2		25 (63%)	15 (37%)
level 1	(lowest in the hierarchical structure)	17 (46%)	20 (54%)

Table 4.9 shows that independent and dependent segments were not distributed equally over the hierarchical levels: the higher in the hierarchical structure, the more frequently independent segments occurred ($\chi^2(3) = 24.56$, $p < .001$). To avoid a confounding effect of it, syntactic class was included in each of the further analyses as an independent factor with two values, ‘independent’ and ‘dependent’. In the original analyses, the factor text type with two values, ‘descriptive’ and ‘argumentative’, was also included as an independent factor. However, no effects of text type were found. Therefore, this factor was removed from the statistical analyses.

4.3.2 Relative approach

4.3.2.1 Procedure of linear adjacency

In the relative method of linear adjacency, the prosodic features of all pairs of adjacent boundaries in the text were examined. In each pair, one boundary was higher in the hierarchical structure than the other. Pairs of boundaries with equal levels were excluded. Figure 4.4 shows an imaginary hierarchical structure of segments 10 to 13 to illustrate the scoring of adjacent boundaries. From the pair of adjacent boundaries between segments 10 and 11, on the one hand, and between segments 11 and 12, on the other hand, the boundary between segments 11 and 12 is the higher one. From the pair of adjacent boundaries between segments 11 and 12, on the one hand, and between segments 12 and 13, on the other hand, the boundary between segments 11 and 12 is the lower one. The scoring of boundaries as higher or lower was based on the average scores of the six analyses. Three variants of the relative method of linear adjacency were applied corresponding with the procedures for scoring hierarchical levels.



Figure 4.4 Adjacent boundaries scored pairwise for hierarchical level

4.3.2.1.1 Top-down scoring

The four texts contained 92 pairs with adjacent boundaries that differed in hierarchical level when scored top-down. Two-way ANOVAs were run to test the effect of hierarchical level on each of the three prosodic parameters. Hierarchical level was included as a within-group factor (two levels: higher, lower) and syntactic combination based on the syntactic classes of the segments following the boundaries of a pair was included as a between-groups factor (four levels: independent-independent, independent-dependent, dependent-independent, dependent-dependent). Separate analyses were run for the three prosodic parameters: pause duration, F0-maximum, and articulation rate. Table 4.10 presents for each syntactic combination the prosodic means of the higher and lower boundaries in standard scores.

Table 4.10 For each syntactic combination, prosodic characteristics (in standard scores) in relation to hierarchical level of adjacent boundaries (scored top-down)

syntactic class of segments following		hierarchical level of boundary	preceding pause	F0-maximum	articulation rate	n
b_{higher}	b_{lower}					
independent	independent	b_{higher}	0.78	0.38	-0.28	43
		b_{lower}	-0.06	0.03	-0.01	
independent	dependent	b_{higher}	0.20	0.04	0.40	22
		b_{lower}	-0.68	-0.21	0.12	
dependent	independent	b_{higher}	0.62	0.19	0.47	20
		b_{lower}	-0.87	-0.54	-0.40	
dependent	dependent	b_{higher}	-0.43	0.01	0.17	7
		b_{lower}	-1.19	-0.57	-0.13	

Note: b_{higher} = highest boundary of the pair; b_{lower} = lowest boundary of the pair

There were effects of hierarchical level ($F(1, 88) = 43.92, p < .001, \eta^2 = .33$) and syntactic combination ($F(3, 88) = 12.19, p < .001, \eta^2 = .29$) on the preceding pause. Pauses preceding higher boundaries were longer (mean: 0.52) than those preceding lower boundaries (mean: -0.47). For the syntactic combination, all pairwise comparisons (Tukey's HSD procedure) were significant, except that between the independent-dependent and the dependent-independent pair. The more dependent segments involved, the shorter the corresponding pauses. It did not matter whether the high or the low boundary was followed by a dependent segment. There was no interaction between hierarchy and syntactic status for the preceding pause ($F(3, 88) = 1.59, p = .20$).

There was an effect of hierarchical level on the F0-maximum ($F(1, 88) = 6.68, p < .05, \eta^2 = .07$). F0-maxima of segments following higher boundaries were higher (mean: 0.23) than F0-maxima of segments following lower boundaries (mean: -0.20). There was no effect of syntactic combination ($F(3, 88) = 2.27, p = .09$), and no interaction between hierarchy and syntactic combination for the F0-maximum ($F < 1$).

There were no effects of hierarchical level ($F(1, 88) = 2.49, p = .12$) and syntactic combination ($F(3, 88) = 2.02, p = .12$) on articulation rate. There was an interaction between hierarchy and syntactic combination ($F(3, 88) = 2.96, p < .05, \eta^2 = .09$). In the independent-independent combination, segments following higher boundaries were read more slowly (fewer phonemes per second) and segments following lower boundaries faster, whereas in the other syntactic combinations this pattern was reversed.

4.3.2.1.2 Bottom-up scoring

The four texts contained 108 pairs with adjacent boundaries that differed in hierarchical level when scored bottom-up. The same statistical analyses were performed as described in section 4.3.2.1.1. Table 4.11 presents for each syntactic combination the prosodic means of the higher and lower boundaries in standard scores.

Table 4.11 For each syntactic combination, prosodic characteristics (in standard scores) in relation to hierarchical level of adjacent boundaries (scored bottom-up)

syntactic class of segments following		hierarchical position of boundary	preceding pause	F0-maximum	articulation rate	n
b_{higher}	b_{lower}					
independent	independent	b_{higher}	0.74	0.32	-0.14	52
		b_{lower}	-0.06	-0.07	-0.08	
independent	dependent	b_{higher}	0.33	0.02	0.54	24
		b_{lower}	-0.74	-0.28	0.01	
dependent	independent	b_{higher}	0.53	0.12	0.50	25
		b_{lower}	-0.88	-0.42	-0.38	
dependent	dependent	b_{higher}	-0.78	0.21	0.18	7
		b_{lower}	-1.20	-0.78	0.09	

Note: b_{higher} = highest boundary of the pair; b_{lower} = lowest boundary of the pair

There were effects of hierarchical level ($F(1, 104) = 43.53, p < .001, \eta^2 = .30$) and syntactic combination ($F(3, 104) = 14.90, p < .001, \eta^2 = .30$) on the preceding pause. Pauses preceding higher boundaries were longer (mean: 0.50) than those preceding lower boundaries (mean: -0.47). For the syntactic combination, all pairwise comparisons (Tukey's HSD procedure) were significant, except that between the independent-dependent and the dependent-independent pair. The more dependent segments involved, the shorter the corresponding pauses. Whether the high or the low boundary was followed by a dependent segment was of no influence. There was no interaction between hierarchy and syntactic status for pause duration ($F(3, 104) = 2.21, p = .09$).

There was an effect of hierarchical level on the F0-maximum ($F(1, 104) = 10.54, p < .05, \eta^2 = .09$). F0-maxima of segments following higher boundaries were higher (mean: 0.20) than F0-maxima of segments following lower boundaries (mean: -0.24). There was no effect of syntactic combination ($F(3, 104) = 1.71, p = .17$) and no interaction between hierarchy and syntactic combination ($F < 1$) for the F0-maximum.

There was an effect of hierarchical level ($F(1, 104) = 4.00, p < .05, \eta^2 = .04$) on articulation rate. Segments following higher boundaries were read faster (mean: 0.18) than segments following lower boundaries (mean: -0.12). There was no effect of syntactic combination ($F(3, 104) = 2.04, p = .11$) and no interaction of hierarchy and syntactic status ($F(3, 104) = 2.58, p = .06$) for articulation rate.

4.3.2.1.3 Symmetrical scoring

The four texts contained 101 pairs with adjacent boundaries that differed in hierarchical level when scored symmetrically. The same statistical analyses were performed as described in section

4.3.2.1.1. Table 4.12 presents for each syntactic combination the prosodic means of the higher and lower boundaries in standard scores.

Table 4.12 For each syntactic combination, prosodic characteristics (in standard scores) in relation to hierarchical level of adjacent boundaries (scored symmetrically)

syntactic class of segments following		hierarchical position of boundary	preceding pause	F0-maximum	articulation rate	n
b_{higher}	b_{lower}					
independent	independent	b_{higher}	0.76	0.36	-0.07	48
		b_{lower}	-0.10	-0.07	-0.11	
independent	dependent	b_{higher}	0.28	0.15	0.45	23
		b_{lower}	-0.61	-0.29	-0.02	
dependent	independent	b_{higher}	0.58	0.15	0.56	24
		b_{lower}	-0.92	-0.44	-0.39	
dependent	dependent	b_{higher}	-0.92	-0.06	-0.08	6
		b_{lower}	-1.38	-0.87	0.16	

Note: b_{higher} = highest boundary of the pair; b_{lower} = lowest boundary of the pair

There were effects of hierarchical level ($F(1, 97) = 35.36, p < .001, \eta^2 = .27$) and syntactic combination ($F(3, 97) = 14.64, p < .001, \eta^2 = .31$) on the preceding pause. Pauses preceding higher boundaries were longer (mean: 0.51) than those preceding lower boundaries (mean: -0.49). For the syntactic combination all pairwise comparisons (Tukey's HSD procedure) were significant, except for that between the independent-dependent and the dependent-independent pair. The more dependent segments involved, the shorter the corresponding pauses. Whether the high or the low boundary was followed by a dependent segment was of no influence. There was no interaction between hierarchy and syntactic status ($F(3, 97) = 2.11, p = .10$).

There was an effect of hierarchical level on the F0-maximum ($F(1, 97) = 9.98, p < .05, \eta^2 = .09$). F0-maxima of segments following higher boundaries were higher (mean: 0.24) than F0-maxima of segments following lower boundaries (mean: -0.25). There was no effect of syntactic combination on the F0-maximum ($F(3, 97) = 2.00, p = .12$), and no interaction between hierarchy and syntactic combination ($F < 1$).

There was no effect of hierarchical level on articulation rate ($F(1, 97) = 2.64, p = .11$). Segments following higher boundaries were read as fast as segments following lower boundaries. There was no effect of syntactic combination on articulation rate ($F(3, 97) = 1.30, p = .28$) and no interaction between hierarchy and syntactic status ($F(3, 97) = 2.42, p = .07$).

4.3.2.2 Procedure of hierarchical adjacency

In the procedure of hierarchical adjacency, it was determined for each pair of boundaries which had a direct relation of superordination or subordination, which one of the two was the higher in

the hierarchical structure. In each pair, one boundary was higher in the hierarchical structure than the other, because pairs of boundaries with equal levels were excluded from the analyses. Figure 4.5 presents an imaginary hierarchical structure of segments 13 to 23. Boundary 19-20 dominates both boundary 14-15 and boundary 22-23, because the boundary between segments 19 and 20 is the boundary between the text span consisting of segments 13 to 19, on the one hand, and the text span consisting of segments 20 to 23, on the other hand, and because the text span consisting of segments 13 to 19 is, one level lower, further divided in the text spans consisting of segments 13 and 14, on the one hand, and segments 15 to 19, on the other hand, and the text span consisting of segments 20 to 23 is, one level lower, further divided in the text span consisting of segments 20 to 22, on the one hand, and segment 23, on the other hand. Therefore, the boundary between segments 19 and 20 is pairwise compared with the boundaries between segments 14 and 15 and between segments 22 and 23 with regard to their prosodic realizations. In the same way, boundary 14-15 dominates both boundary 13-14 and boundary 16-17. Therefore, the boundary between segments 14 and 15 is pairwise compared with the boundaries between segments 13 and 14 and between segments 16 and 17 with regard to their prosodic realizations.

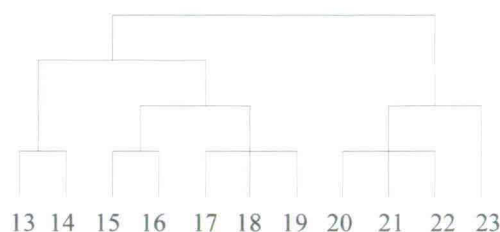


Figure 4.5 Dominating boundaries scored pairwise for hierarchical level

For each text, the scoring of the boundaries as higher or lower was based on the level scores of the RST analysis which gave the highest mean pairwise correlation between the six analysts. In the procedure of hierarchical adjacency, only the top-down procedure of scoring the boundary levels was applied, and related to the prosodic realizations of the segments following the boundaries. The bottom-up and symmetrical scorings were not applied anymore since the effects on prosody found in the separate analyses of the procedure of linear adjacency hardly differed. The top-down scoring was chosen, because it was associated most with the hierarchical representation in terms of the distance between the boundary levels and the top level of the hierarchy.

The four texts contained 84 pairs of boundaries which had a direct relation of superordination or subordination, when scored top-down. Two-way ANOVAs were run to test the effect of hierarchical level on each of the three prosodic parameters. Hierarchical level was included as a within-group factor (two levels: higher, lower) and syntactic combination based on the syntactic classes of the segments in a pair was included as a between-groups factor (four levels:

independent-independent, independent-dependent, dependent-independent, dependent-dependent). Separate analyses were run for the three prosodic parameters: pause duration, F0-maximum, and articulation rate. Table 4.13 presents for each syntactic combination the prosodic means of the higher and lower boundaries in standard scores.

Table 4.13 For each syntactic combination, prosodic characteristics (in standard scores) in relation to hierarchical level of dominating boundaries

syntactic class of segments following		hierarchical position of boundary	preceding pause	F0-maximum	articulation rate	n
b_{higher}	b_{lower}					
independent	independent	b_{higher}	0.61	0.48	-0.09	51
		b_{lower}	0.23	0.18	-0.02	
independent	dependent	b_{higher}	0.29	-0.01	0.31	23
		b_{lower}	-0.77	-0.13	-0.14	
dependent	independent	b_{higher}	-0.49	-0.17	0.06	5
		b_{lower}	0.17	0.07	0.81	
dependent	dependent	b_{higher}	-1.10	0.46	0.48	5
		b_{lower}	-1.14	-1.00	0.04	

Note: b_{higher} = highest boundary of the pair; b_{lower} = lowest boundary of the pair

There was no effect of hierarchical level on pause duration ($F < 1$). Syntactic combination affected pause duration ($F(3, 80) = 19.41, p < .001, \eta^2 = .42$). For syntactic combination, all pairwise comparisons (Tukey's HSD procedure) were significant, except the comparison between the independent-dependent combination and the dependent-independent combination. There was an interaction between hierarchy and syntactic combination for the preceding pause ($F(3, 80) = 3.55, p < .05, \eta^2 = .12$). The differences in pause durations between higher and lower boundaries differed for the four syntactic combinations. The difference between higher and lower boundaries for pauses was greatest in the independent-dependent combination; it was least in the dependent-dependent combination. The pattern that segments following higher boundaries have longer preceding pauses than segments following lower boundaries, was reversed in the dependent-independent combination.

There was an effect of hierarchical level on the F0-maximum ($F(1, 80) = 4.52, p < .05, \eta^2 = .05$). The F0-maximum of segments following higher boundaries was higher (mean: 0.31) than that of segments following lower boundaries (mean: -0.02). There was no effect of syntactic combination ($F(3, 80) = 1.90, p = .14$). There was no interaction between hierarchy and syntactic combination ($F(3, 80) = 2.16, p = .19$).

There were no effects of hierarchical level ($F < 1$) and syntactic combination ($F(3, 80) = 2.37, p = .08$) on articulation rate. There was no interaction between hierarchy and syntactic combination for articulation rate ($F(3, 80) = 1.07, p = .37$).

4.3.2.3 Summary of results

Table 4.14 presents for the procedures of the relative approach an overview of the effects of hierarchy and syntactic combination and their interactions on the three prosodic parameters. Not surprisingly, the results of the three linear procedures were similar (see Table 4.6), except that in only one of the procedures an effect on articulation rate was found. For all three variants of the procedure, an effect was found of hierarchy on pause duration and on the F0-maximum. Pauses of higher boundaries were longer than pauses of lower boundaries, and the F0-maxima of segments following higher boundaries were higher than those of segments following lower boundaries. In the three variants of the procedure of linear adjacency, no clear pattern emerged for articulation rate. In the procedure of hierarchical adjacency, hierarchy did not affect pause duration, but it affected the F0-maximum. All analyses showed that the syntactic class of the segments of the pairs is a factor to take into consideration.

Table 4.14 For the procedures of the relative approach, an overview of the effects of hierarchy and syntactic combination and their interactions on prosodic characteristics

		preceding pause	F0-maximum	articulation rate
Linear adjacency				
top-down	hierarchy	$p < .001$	$p < .05$	-
	syntactic combination	$p < .001$	-	-
	hierarchy x syntactic combination	-	-	$p < .05$
bottom-up	hierarchy	$p < .001$	$p < .05$	$p < .05$
	syntactic combination	$p < .001$	-	-
	hierarchy x syntactic combination	-	-	-
symmetrical	hierarchy	$p < .001$	$p < .05$	-
	syntactic combination	$p < .001$	-	-
	hierarchy x syntactic combination	-	-	-
Hierarchical adjacency				
	hierarchy	-	$p < .05$	-
	syntactic combination	$p < .001$	-	-
	hierarchy x syntactic combination	$p < .05$	-	-

4.3.3 Absolute approach

In the absolute approach, the prosodic characteristics were evaluated directly in relation to the absolute scores of the levels, defined on an interval scale. The three procedures were performed to score the levels. The syntactic classes of the segments were included in the statistical analyses.

4.3.3.1 Top-down scoring

For the four texts taken together, the top-down scores of the levels of the boundaries averaged over the six RST analyses were evaluated in relation to the pause durations of those boundaries, the F0-maxima, and the articulation rates of the segments following those boundaries. Table 4.15 presents for each hierarchical level the mean standard scores of pause duration, F0-maximum, and articulation rate. It also presents the percentages of independent segments per level in order to show the distribution of independent and dependent segments over the levels. There are three level-1 observations for the four texts taken together, whereas four might have been expected, since the level scores were averaged over the six RST analyses: in one case the rounding-off changed a score 1 into a score 2.

Table 4.15 Prosodic characteristics for each hierarchical level scored top-down (level 1 = highest boundary)

		independent segments	preceding pause	F0-maximum	articulation rate
level 1	(n=3)	100 %	0.97	0.51	0.51
level 2	(n=7)	71 %	0.38	-0.14	-0.45
level 3	(n=17)	82 %	0.61	0.11	-0.02
level 4	(n=16)	88 %	0.26	0.29	-0.30
level 5	(n=23)	83 %	0.07	0.13	-0.05
level 6	(n=24)	50 %	-0.41	-0.25	0.33
level 7	(n=11)	64 %	-0.03	-0.12	0.58
level 8	(n=11)	36 %	-0.81	-0.06	-0.33
level 9	(n=2)	100 %	0.26	-0.22	-0.61
level 10	(n=6)	50 %	-0.34	-0.55	-0.03
level 11	(n=1)	100 %	-0.95	0.74	0.32

Table 4.15 shows roughly that, as the hierarchical level of the boundaries decreased, the durations of the pauses of those boundaries and the F0-maxima of the segments following those boundaries also decreased. There was no such pattern for articulation rate. Note that there were little observations at the highest and the lowest levels. Therefore, we formed new classes of levels. Levels 1 and 2 were taken together (and called level 1), and the same was done to levels 3 and 4 (called level 2), levels 5 and 6 (called level 3), and levels 7 to 11 (called level 4).

Two-way analyses of variance were run with hierarchical level and syntactic class as the independent factors and each of the three prosodic parameters as the dependent factor. Table 4.16 presents for each hierarchical class the mean standard scores of the three prosodic parameters for independent and dependent segments separately.

Table 4.16 Prosodic characteristics for each hierarchical class scored top-down for independent and dependent segments (level 1 = highest boundary)

			preceding pause	F0-maximum	articulation rate
independent segments	level 1	(n=8)	0.85	0.51	0.07
	level 2	(n=28)	0.62	0.31	-0.15
	level 3	(n=31)	0.13	0.06	0.21
	level 4	(n=17)	0.15	-0.04	0.16
dependent segments	level 1	(n=2)	-0.64	-1.73	-1.09
	level 2	(n=5)	-0.52	-0.37	-0.16
	level 3	(n=16)	-0.76	-0.29	0.03
	level 4	(n=14)	-1.02	-0.32	-0.06

Pause duration was not affected by hierarchy ($F(3, 113) = 1.82, p=.15$), but it was by syntactic class ($F(1, 113) = 30.33, p<.001, \eta^2 = .21$). Pauses preceding independent segments were longer than pauses preceding dependent segments (0.36 versus -0.81). There was no interaction between hierarchy and syntactic class ($F<1$).

The F0-maximum was not affected by hierarchy ($F<1$), but it was by syntactic class ($F(1, 113) = 12.43, p<.001, \eta^2 = .09$). The F0-maximum of independent segments was higher than that of dependent segments (0.16 versus -0.39). There was no interaction between hierarchy and syntactic class for the F0-maximum ($F(3, 113) = 2.01, p=.12$).

There were no effects of hierarchy and syntactic class on articulation rate, and no interaction (hierarchy: $F<1$; syntactic class: ($F(1, 113) = 2.28, p=.13$; hierarchy x syntactic class: $F<1$).

Since the hypothesis was that as hierarchical level decreases, pause duration and the F0-maximum also decrease and articulation rate increases, linear trends were computed with the hierarchical classes for the three prosodic parameters. For the independent segments, the linear trend was significant for pause duration ($F(1, 80) = 7.28, p<.01$), but not for the F0-maximum ($F(1, 80) = 2.53, p=.12$) and articulation rate ($F<1$). For the dependent segments, none of the linear trends was significant (pause duration: $F(1, 33) = 1.74, p=.20$; F0-maximum: $F(1, 33) = 1.80, p=.19$; articulation rate: $F<1$).

4.3.3.2 Bottom-up scoring

For the four texts together, the bottom-up scores of the levels of the boundaries averaged over the six RST analyses were evaluated in relation to the pause durations of these boundaries, and the F0-maxima and the articulation rates of the segments following these boundaries. Table 4.17 presents for each hierarchical level mean standard scores of pause duration, F0-maximum, and articulation rate, and the percentages of independent segments.

Table 4.17 Prosodic characteristics for each hierarchical level scored bottom-up (level 1 = lowest boundary)

		independent segments	pause duration	F0-maximum	articulation rate
level 1	(n = 35)	43 %	-0.74	-0.28	-0.25
level 2	(n = 28)	64 %	-0.10	-0.20	0.13
level 3	(n = 15)	67 %	-0.02	-0.34	0.16
level 4	(n = 15)	87 %	0.30	0.35	0.09
level 5	(n = 8)	100 %	1.11	0.90	-0.34
level 6	(n = 11)	100 %	0.88	0.24	0.63
level 7	(n = 4)	100 %	0.22	0.81	-0.57
level 8	(n = 1)	100 %	-0.20	-1.14	1.30
level 10*	(n = 3)	100 %	0.82	0.74	0.15
level 11	(n = 1)	100 %	2.79	0.43	-0.09

* Note: Level 9 did not occur as an average score of the six RST analyses when scored bottom-up

There were few data at the highest levels. Therefore, we formed new classes of levels. Levels were taken together: levels 2 and 3 (called level 2), levels 4 to 6 (called level 3), and levels 7 to 11 (called level 4). The distribution over the resulting four classes was similar to the distribution over the classes in the top-down scoring: the class of highest boundaries contained about 10 observations and all other classes about 35 observations. For the independent segments, a one-way analysis of variance was run with hierarchy (four levels) as the independent factor, and each of the three prosodic parameters as the dependent factors. The dependent segments were not analyzed, because of the low number of observations at level 3, and no observations at level 4. Table 4.18 presents for each hierarchical class mean standard scores of the three prosodic parameters for independent and dependent segments separately.

Table 4.18 Prosodic characteristics per hierarchical class scored bottom-up for independent and dependent segments (level 1 = lowest boundary)

			pause duration	F0-maximum	articulation rate
independent segments	level 1	(n=15)	-0.38	0.03	-0.13
	level 2	(n=28)	0.24	-0.21	0.10
	level 3	(n=32)	0.73	0.44	0.16
	level 4	(n=9)	0.66	0.53	-0.07
dependent segments	level 1	(n=20)	-1.01	-0.52	-0.35
	level 2	(n=15)	-0.65	-0.33	0.22
	level 3	(n=2)	-0.12	0.46	0.13

* Note: Level 4 did not occur for dependent segment since the original boundary levels 7 to 11 were only followed by independent segments

For the independent segments, pause duration was affected by hierarchy ($F(3, 80) = 8.09, p < .001, \eta^2 = .23$). The F0-maximum was affected by hierarchy ($F(3, 80) = 2.93, p < .05, \eta^2 = .10$). The articulation rate was not affected by hierarchy ($F < 1$).

Linear trends were computed on the hierarchical classes for the three prosodic parameters. For independent segments, the linear trends were significant for pause duration ($F(1, 80) = 20.55, p < .001$) and the F0-maximum ($F(1, 80) = 4.98, p < .05$), but not for articulation rate ($F < 1$). None of the linear trends was significant for the dependent segments (pause duration: $F(1, 33) = 3.87, p = .06$; F0-maximum: $F(1, 33) = 1.65, p = .21$; articulation rate: $F(1, 33) = 2.62, p = .12$). Although the linear trend was not significant for pause duration and the F0-maximum for the dependent segments, it should be noted that there were only two observations for level 3, and there were no observations for level 4. The means of levels 1 and 2 showed the same trend as for the independent segments.

4.3.3.3 Symmetrical scoring

For the four texts taken together, the symmetrical scores of the levels of the boundaries averaged over the six RST analyses were related to the pause durations of these boundaries, and the F0-maxima and the articulation rates of the segments following these boundaries. Table 4.19 presents mean standard scores of pause duration, F0-maximum, and articulation rate per hierarchical level, and the percentages of independent segments per level.

Table 4.19 Prosodic characteristics per hierarchical level scored symmetrically (level 1 = lowest boundary)

		independent segments	pause duration	F0-maximum	articulation rate
level 1	(n=37)	46 %	-0.68	-0.28	-0.24
level 2	(n=40)	63 %	-0.11	-0.21	0.09
level 3	(n=19)	95 %	0.18	0.11	-0.01
level 4	(n=11)	100 %	0.93	0.67	0.64
level 5	(n=8)	88 %	0.87	0.35	-0.26
level 6	(n=5)	100 %	1.24	0.96	0.48
level 9*	(n=1)	100 %	2.79	0.43	-0.09

* Note: Levels 7 and 8 did not occur as average scores of the six RST analyses when scored symmetrically

There were few boundaries with high-level scores. Therefore, we formed new classes of levels. Levels were taken together: levels 3 and 4 were called level 3, and levels 5 to 9 were called level 4. The distribution of the classes was similar to that of the top-down and bottom-up scoring: the class of high boundaries contained about 10 observations and all other classes about 35. Two-way analyses of variance were run with hierarchy and syntactic class as independent factors, and each of the three prosodic parameters as dependent factors. Table 4.20 presents for each hierarchical

class mean standard scores of pause duration, F0-maximum, and articulation rate for independent and dependent segments separately.

Table 4.20 Prosodic characteristics per hierarchical class scored symmetrically for independent and dependent segments (level 1 = lowest boundary)

			pause duration	F0-maximum	articulation rate
independent segments	level 1	(n=17)	-0.28	0.01	-0.12
	level 2	(n=25)	0.15	-0.13	-0.05
	level 3	(n=29)	0.55	0.31	0.28
	level 4	(n=13)	1.17	0.58	0.05
dependent segments	level 1	(n=20)	-1.01	-0.52	-0.35
	level 2	(n=15)	-0.55	-0.33	0.33
	level 3	(n=1)	-2.43	0.41	-1.14
	level 4	(n=1)	0.72	0.58	-0.36

Pause duration was affected by hierarchy ($F(3, 113) = 7.68, p < .001, \eta^2 = .17$) and by syntactic class ($F(1, 113) = 20.13, p < .001, \eta^2 = .15$). The higher the level in the hierarchy of the segments, the longer the preceding pauses, and pauses were longer when they preceded independent segments than when they preceded dependent segments. The pattern was difficult to characterize for pauses preceding dependent segments, because of the low number of observations at levels 3 and 4. There was an interaction between hierarchy and syntactic class ($F(3, 113) = 3.20, p < .05, \eta^2 = .08$).

The F0-maximum was not affected by hierarchy ($F(3, 113) = 1.32, p = .27$), or by syntactic class ($F < 1$). There was no interaction ($F < 1$).

Articulation rate was not affected by hierarchy ($F(3, 113) = 1.08, p = .36$), or by syntactic class ($F(1, 113) = 1.23, p = .27$). There was no interaction ($F(3, 113) = 1.37, p = .26$).

Linear trends were computed on the hierarchical classes for the three prosodic parameters. For independent segments, the linear trends were significant for pause duration ($F(1, 80) = 33.83, p < .001$) and the F0-maximum ($F(1, 80) = 4.33, p < .05$), but not for articulation rate ($F < 1$). None of the linear trends was significant for the dependent segments (pause duration: $F(1, 33) = 2.99, p = .09$; F0-maximum: $F(1, 33) = 1.98, p = .17$; articulation rate: $F < 1$). Although the linear trend was not significant for pause duration and the F0-maximum for the dependent segments, it should be noted that the categories of level 3 and 4 each contained only one observation. The means of levels 1 and 2 showed the same trend as for the independent segments.

4.3.3.4 Summary of results

Table 4.21 presents for the procedures of the absolute approach an overview of the effects of hierarchy, syntactic class, and the interaction between these two factors for the three prosodic parameters. The results of the linear trends are also presented.

The linear trends for pause duration were significant in the three procedures for the independent segments. For the three procedures, this meant the lower the boundaries in the hierarchical structure, the shorter the pause durations. This pattern of pause durations makes clear that speakers do not simply divide a text into successive paragraphs. While varying their pause durations, they appear to produce a more subtle equivalent of the hierarchical structure of the entire text.

An effect of hierarchy on the F0-maximum was found in the linear trends in both the bottom-up and the symmetrical procedure for the independent segments. For both procedures, this meant that the lower the segments in the text structure, the lower their F0-maxima. A linear trend was not found for articulation rate in any of the analyses.

Table 4.21 For the procedures of the absolute approach, an overview of the effects of hierarchy and syntactic class and their interaction on prosodic characteristics

		pause	F0-maximum	rate
top-down	hierarchy	-	-	-
	syntactic class	$p < .001$	$p < .001$	-
	hierarchy x syntactic class	-	-	-
	linear trend	$p < .01$	-	-
	independent segments	-	-	-
bottom-up	hierarchy	$p < .001$	$p < .05$	-
	syntactic class	n.a.	n.a.	n.a.
	hierarchy x syntactic class	n.a.	n.a.	n.a.
	linear trend	$p < .001$	$p < .05$	-
	independent segments	-	-	-
symmetrical	hierarchy	$p < .001$	-	-
	syntactic class	$p < .001$	-	-
	hierarchy x syntactic class	$p < .05$	-	-
	linear trend	$p < .001$	$p < .05$	-
	independent segments	-	-	-

Note: n.a. = not applied

Finally, to test the robustness of this study, multiple regressions were performed with the absolute hierarchical levels as the dependent variable and the three prosodic parameters and syntactic class as independent variables. In the top-down procedure, this analysis resulted in a multiple correlation of .39. Only the contribution of pause duration was significant ($p < .01$; F0-maximum: $p = .90$; articulation rate: $p = .12$; syntactic class: $p = .54$). In the bottom-up procedure, the analysis resulted in a multiple correlation of .60. The contributions of pause duration ($p < .001$) and syntactic class ($p < .05$) were significant: those of the other prosodic parameters were not (F0-maximum: $p = .10$; articulation rate: $p = .39$). In the symmetrical procedure, the analysis resulted in a multiple correlation of .62. Only the contribution of pause duration was significant ($p < .001$; F0-maximum: $p = .15$; articulation rate: $p = .30$; syntactic class: $p = .17$). In the three scoring

procedures, the variation in pause duration added to the prediction of the levels in the hierarchical structure; the variation in the F0-maximum was not a significant predictor of hierarchical levels.

4.4 Conclusion and discussion

The hierarchical structure of a text and prosody are related. Pause duration was found to be a strong indicator of the multilayered hierarchical structure of a text.

The results for the F0-maximum were almost as clear as for the pause durations. For the relative approach, effects of hierarchy on the F0-maximum were found in all four procedures. For the absolute approach, the linear trends were significant in the bottom-up and symmetrical procedures for the independent segments. Articulation rate was not affected by the hierarchical levels in any of the procedures, except in the bottom-up procedure of linear adjacency.

A small corpus of natural texts was used in this study, because of the explorative character of it. In the study reported in the following chapter more text and speech material is used to clarify the relation between text structure and prosody in more detail. Based on the results of this study, decisions have to be made with regard to the problem of scoring text structure and the use of various procedures, i.e., relative and absolute procedures.

Three procedures of scoring the hierarchical levels were proposed: top-down, bottom-up, and symmetrical procedures, since there was a possibility that they were related to prosody in different ways. In general, the results of the three methods turned out to be similar. For reasons of effectiveness, one of the three procedures has to be selected for the research reported in Chapter 5. The overview in Table 4.14 shows that the relation between hierarchy and prosody in the relative approach was very similar in the three procedures, except the result with regard to articulation rate. The overview in Table 4.21 shows that the relation between hierarchy and prosody in the absolute approach was slightly stronger in the symmetrical procedure. The bottom-up and symmetrical procedures resembled each other more than the top-down and symmetrical procedures (see Table 4.6). For these two reasons, the top-down procedure is dropped and the symmetrical procedure is preferred to the bottom-up one. The symmetrical procedure will be used in the study described in Chapter 5. After all, the problem of scoring the hierarchical levels has not been solved yet, but the consequences of each procedure have been faced up to.

An absolute and a relative approach were performed since there was a possibility that local and global aspects of text structure would be reflected differently by prosody. Global aspects of hierarchy were thought to be captured better using the absolute approach, whereas local aspects of hierarchy were thought to be captured better using the relative approach. For the absolute approach, it was assumed that speakers have a concept of the hierarchical structure of a text as a whole in their minds, supposing that they have prepared the text before reading it aloud. Their prosodic realizations of the segments and the boundaries between the segments may then be regarded as reflections of their concept of the global hierarchical structure of the text. The assumptions underlying the relative approach may be considered weaker than the assumptions underlying the absolute approach. In the relative approach, the prosodic realizations of the

speakers may be regarded as reflections of their concept of the local hierarchical relations within the text.

For the relative approach of hierarchical adjacency, it was assumed that speakers have a concept of the hierarchical structure within a branch of a text structure in their minds, supposing that they have prepared the text. For the relative approach of linear adjacency, it was assumed that speakers build up their concept of the hierarchical structure of a text incrementally, even if they prepared the text before reading it aloud. In practice, the results of the various approaches hardly differed. Nevertheless, they will be applied in the study reported in the following chapter.

It was not examined in this study whether the four speakers differed with regard to the way they realized the hierarchical structures of their texts. It is possible that some speakers realized the hierarchical structure of the text in a more pronounced way than others, for example, because they were in some way more aware of the hierarchical structure of the text, or because their ways of articulation were more expressive. Especially professional speakers such as those used in this study might be affected by mannerisms in their articulation. Alternatively, the different speakers may have used different prosodic parameters to express the hierarchical structure of the text. Some speakers might have realized the hierarchical structure using pause duration, while others realized it using F0-maximum, articulation rate, or other prosodic parameters. As a result of individual differences between speakers having been ignored, possible effects of the F0-maximum or articulation rate, or both may have disappeared. In the study described in Chapter 5, non-professional speakers will be selected, and specific attention will be given to the individual differences between the speakers.

The role of syntactic status was discussed in section 4.3.1. The syntactic class of the segments was found to be an important factor with respect to prosody. Therefore, in the study reported in the next chapter, syntactic class will be included in all analyses.

The texts used in this study were two descriptive, narrative texts and two argumentative texts. The factor text type was not controlled for. However, text types of a different kind may result in different hierarchical elaboration. In comparison with the hierarchical organization of argumentative texts, the structures of descriptive texts may have a far more linear organization. In the study described in the next chapter, the focus will be on text material of one text type, i.e., descriptive texts.

5

Prosody of hierarchy, nuclearity, and rhetorical relations: A corpus-based study

5.1 Introduction¹

In this study prosody is again evaluated in relation to hierarchy, as well as in relation to other aspects of text structure, namely, nuclearity and rhetorical relations. The central questions are how the various hierarchical levels of the boundaries, the nuclearity of the segments, and the particular rhetorical relations between segments are reflected in prosody. A text corpus of considerable size is built: twenty long descriptive texts are selected. Each of the texts is read aloud by a different speaker. Pause durations preceding the segments, F0-maxima, and the articulation rates of the segments are measured. The texts were analyzed using Rhetorical Structure Theory. The hierarchical levels of the boundaries are scored using the symmetrical procedure. The syntactic classes of the segments are included as a separate factor in the statistical analyses. With respect to hierarchy, the hypothesis is that the higher segments are in the hierarchical structure, the more strongly they will be marked prosodically. For the prosodic parameters, we expect that pauses preceding segments will be longer, and the F0-maxima of segments higher. For nuclearity, the hypothesis is that nuclei will be marked prosodically more strongly than satellites. We expect longer pause durations preceding nuclei, higher F0-maxima in nuclear segments, and a slower articulation rate of nuclear segments. No specific hypotheses are formulated for the prosodic realizations of distinct groups of rhetorical relations.

5.2 Method

5.2.1 Text material

Twenty news reports were selected from a Dutch national quality newspaper. The required length of a text was set at at least fifteen sentences in order to get the complex hierarchical structures necessary for this study. The themes of the reports varied: politics, accidents, crimes, sports, social phenomena. The style of the news reports was objective and non-controversial. Slight syntactic changes were made in some formulations to facilitate the segmentation process. Most changes concerned direct speech that was changed into indirect speech. They were few in number. Table 5.1 presents one of the texts as an example (the original Dutch text is presented in Appendix C). The texts were divided into basic segments (clauses) according to the criteria given by RST. Problematic cases were those parts of sentences which formally had to be considered separate segments, although they could be understood only as parts of other segments. In general, these parts of sentences were considered part of their host sentences.

The mean length of the texts was 28 segments, with a range from 19 to 37. After segmentation, the twenty texts were analyzed using RST with respect to hierarchical levels, nuclei and satellites, and rhetorical relations. The analyses were verified by a second trained user of RST. Based on the reliability studies mentioned earlier, it was assumed that these twenty text analyses could be considered plausible interpretations of the texts. The three text-structural aspects investigated, hierarchy, nuclearity, and rhetorical relations, are explained in the following section.

¹ An earlier version of this chapter was published in Den Ouden, Noordman & Terken (2003).

5.2.2 Text-structural characteristics

5.2.2.1 Hierarchy

To investigate the relation between the levels of the hierarchical structure of the texts and the prosodic realization of the segments following the boundaries at those levels, each boundary between segments was given a score. Figure 5.1 presents the RST analysis of the sample text in Table 5.1. The scores were assigned to the levels according to the symmetrical procedure described in Chapter 4. Figure 5.2 presents the bottom-up representation of the text analysis in Figure 5.1.

The twenty texts contained a total of 543 boundaries. The level scores ranged from 1 to 10. In the statistical analyses these ten levels were reduced to five classes, because there were few boundaries with a score of 4 or 5, and even fewer boundaries with a score of 6 or higher. Therefore, scores 4 and 5 were put together into one class, as were scores 6 to 10. This resulted in 210 boundaries for level 1, 134 for level 2, 76 for level 3, 77 for level 4, and 46 for level 5.

5.2.2.2 Nuclearity

In RST, segments are analyzed as either nuclei or satellites. In the sample text (Table 5.1), the nuclei were segments 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 17, and 18 to 25; and the satellites segments 2, 5, 8, 11, 14, 16, 26, and 27 (see Figure 5.1). The twenty texts together contained 383 nuclei and 180 satellites. Nuclei outnumbered satellites, because many segments were connected by a Joint, Sequence, or Contrast relation. These rhetorical relations consisted of two (or more) nuclei.

5.2.2.3 Rhetorical relations

To investigate the relation between rhetorical relations and prosody, the boundaries in the text structures were classified in accordance with the associated rhetorical relations in the following way. The relation between segments 1 and 2 in Figure 5.1, for example, is characterized by an Elaboration, therefore, the boundary between segments 1 and 2 is classified as Elaboration; the relation between segments 4 and 5 is characterized by Background, therefore, the boundary between segments 4 and 5 is classified as Background. Strictly speaking, the Background relation does not exist between segments 4 and 5, but it exists between segments 3 and 4, on the one hand, and segment 5, on the other hand. The relation name is still attributed to the boundary between segments 4 and 5, and segment 5 is considered the second segment of this Background relation.

It is explored whether the various rhetorical relations differ with respect to the pause duration of the boundaries, the F0-maximum, and the articulation rate of the segments following the boundaries. Twenty-one rhetorical relations occurred in the text structures: Elaboration (n=119), Joint (n=108), Background (n=48), Cause (n=39), Result (n=36), Concession (n=26), Sequence (n=25), Contrast (n=22), Circumstance (n=21), Interpretation (n=20), Restatement (n=17), Antithesis (n=15), Evaluation (n=14), Justify (n=8), Condition (n=7), Solutionhood (n=6), Purpose (n=6), Motivation (n=2), Enablement (n=2), Evidence (n=1), and Summary (n=1). Only

rhetorical relations that occurred more than ten times were included in the statistical analyses. This was the case for 13 rhetorical relations.

A distinction is often made in the literature between causal and non-causal relations (Sanders, Spooren, Noordman, 1992). Segments may be connected either weakly (additively) or strongly (causally). An additive relation exists if only a conjunction relation can be deduced between two segments. A causal relation exists if a relevant implication relation can be deduced. Based on the finding of Sanders and Noordman (2000) that causal relations are comprehended faster than non-causal relations, the question has been raised whether the prosody of causal relations differs from the prosody of non-causal relations. The rhetorical relations in our material were classified as causal and non-causal relations. The causal relations that occurred more than ten times were Cause, Result and Concession. The non-causal relations that occurred more than ten times were Elaboration, Joint, Background, Sequence, Contrast, Circumstance, Interpretation, Restatement, Antithesis, and Evaluation.

Another distinction frequently made in the literature is that between semantic and pragmatic relations (Sweetser, 1990; Sanders, Spooren, Noordman, 1992). A rhetorical relation is semantic if the coherence between the segments in the text is based on the coherence between the events in the world which are described. The segments are considered to cohere because there is a relation between the events in the world. A rhetorical relation is pragmatic if the coherence between the segments in the text is based on the illocutionary meaning of one or both of the segments, for example, when a writer or speaker draws a conclusion. The segments are considered to cohere because of the thought process of the writer or speaker. With regard to prosody, Sweetser (1990: 82) argued that pragmatic readings of a pair of segments require a comma, whereas semantic readings do not. For example, between the semantically related segments '(s1) Anna loves Victor (s2) because he reminds her of her first love' there is a consequence-cause relation of two events in the world. According to Sweetser, the pair is read without a comma, which indicates that the consequence in the first segment is presupposed and that only the causal relation between both segments is affirmed. However, in the pragmatic pair '(s1) Anna loves Victor, (s2) because she told me so herself, and besides, she'd never have proofread his thesis otherwise' ('I conclude that she loves him because I know the relevant data'), the conclusion in the first segment can not be presupposed, and for that reason the pair is read with comma-intonation. Sweetser did not investigate her claims empirically. In this study, the question was raised whether the prosody of pragmatic relations differs from the prosody of semantic relations. The rhetorical relations in our material were classified as semantic and pragmatic relations based on the classification of Mann and Thompson (1988), who referred to 'subject matter' and 'presentational' relations. Semantic or 'subject matter' relations that occurred more than ten times were Elaboration, Circumstance, Cause, Result, Interpretation, Evaluation, Restatement, Sequence, and Contrast. Pragmatic or 'presentational' relations that occurred more than ten times were Antithesis, Background, and Concession. Because the Joint relation could not be classified either as semantic or as pragmatic, it was excluded from the analyses.

Table 5.1 Sample text

	segment
1	The census in China has been extended by five days.
2	Normally it should have ended last Friday.
3	However, millions of people avoided the pollsters
4	or refused to open their doors.
5	This boycott was intended to keep secret illegal children or addresses.
6	During emergency talks last Friday, the government, i.e., the Chinese cabinet, decided to extend the census.
7	An official of the census committee in Beijing said that the committee's employees noticed that it was rather difficult to find active people at home by day or during the evening, but that of course many other people avoided them deliberately
8	At least eighty million farmers, and that number could even be two hundred million, have squatted in the cities.
9	They had themselves registered at their addresses of origin
10	or they were not counted at all.
11	Although they were officially assured that the census has nothing to do with the police,
12	many people are afraid of reprisals when it is discovered that they do not have residence permits.
13	Also many married people who have more than one child boycotted the census,
14	because they were afraid that the committee of birth control would find this out.
15	The employees of the demographic committee now admit that most people did not keep the one-child policy.
16	One of these days even a family with ten children has been found in the region Shanxi.
17	In the opinion of the authorities, the counting of the homeless is not problematic.
18	It was said that there would not be many homeless people
19	and most of them would have an address in another region.
20	They would be counted at these addresses.
21	However, how this counting should happen is unclear.
22	Other people argue that their privacy is affected.
23	These people were found especially in the random sample of ten percent of the population which had to answer 49 detailed questions.
24	The remaining 90 percent only had to answer nineteen general questions
25	People who have not yet been counted are encouraged by advertisements to report to the census committee.
26	The people who have avoided meeting the pollsters thus far will probably not answer this call.
27	It has not proved helpful to the accuracy of this fifth census in the 51-year-old history of the People's Republic of China.

Source: *de Volkskrant*, 3 November 2000

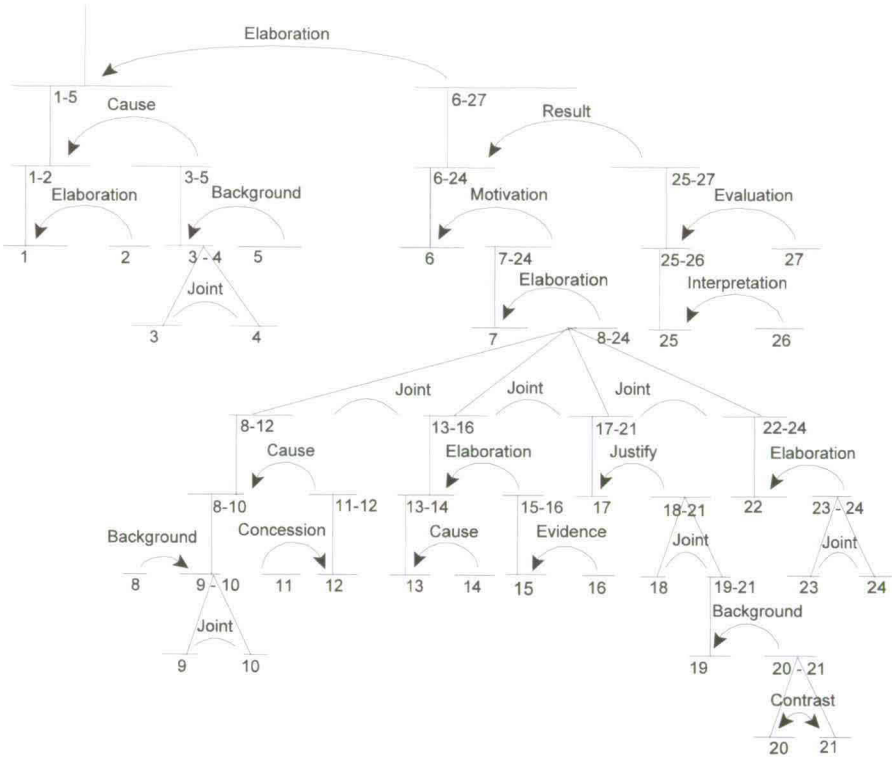


Figure 5.1 RST analysis of the sample text in Table 5.1



Figure 5.2 Bottom-up representation of the RST analysis in Figure 5.1

5.2.3 Procedure

The twenty written news reports were presented to twenty native speakers of Standard Dutch, ten males and ten females, most of them advanced students or employees at the Center for User-System Interaction at Technische Universiteit Eindhoven and the Faculty of Arts at Tilburg University. They were highly educated people with much reading experience. Each speaker read aloud one text. The texts were presented without paragraph markers, but they contained capitals and punctuation marks. The speakers prepared the reading session carefully. They were instructed to imagine that they had to read the news reports for a blind person as clearly as possible. They were encouraged to make notes in the text to improve the reading-aloud task. The preparation of the reading-aloud task was intended to focus the readers' attention on the structure of the text and to enable them to read it aloud as much as possible in accordance with their mental representations of the text structure. The recordings were made in a sound-proof room using a DAT-recorder. The speech was digitized using the speech-processing program Gipos.

5.2.4 Speech material

The beginnings and endings of the pauses of the boundaries between segments were marked in the speech material by hand. Next, the durations were determined automatically in milliseconds.

Pitch range, operationalized as the F0-maximum, was measured for each segment automatically in hertz (Den Ouden & Terken, 2001). The pitch contour of each individual segment was inspected for pitch-measurement errors using an analysis-resynthesis procedure. Pitch-measurement errors consisted mostly of voiced-unvoiced errors and incorrect outliers in the speech signal. These errors were corrected in the speech signal by hand before the automatic measurements procedure was applied. F0-maxima associated with final rises were also removed before the automatic measurement procedure was applied².

Articulation rate was defined as the number of phonemes per second. The number of phonemes in a segment was computed automatically on the hand-corrected canonical transcriptions of the segments using a program called SampaCount.³

Table 5.2 presents the prosodic characteristics of the twenty texts for both female and male speakers. *Pause durations of 0 milliseconds occurred in the speech material, meaning that some sequences of RST-defined segments were read aloud without intervening pauses.* The minimum F0-maximum in Table 5.2 refers to the lowest value of the F0-maximum in the speech material. The number of scores for pause duration is smaller than that for the F0-maximum and articulation rate, because no pauses preceded the first segment of a text.

² Thanks to Leo Vogten for programming automatic means to measure F0-maxima (Department of Electrical Engineering, Technische Universiteit Eindhoven, The Netherlands)

³ Thanks to Jan Roelof de Pijper for programming SampaCount (Department of Technology Management, Technische Universiteit Eindhoven, The Netherlands)

Table 5.2 Prosodic characteristics of the twenty news reports for female and male speakers

			minimum	maximum	mean	standard deviation
pause duration (in milliseconds)	female	(n=268)	0	2298	801	374
	male	(n=275)	0	2380	917	426
F0-maximum (in hertz)	female	(n=278)	194	364	281	34
	male	(n=285)	99	252	169	29
articulation rate (in phonemes per second)	female	(n=278)	10.5	18.4	14.6	1.50
	male	(n=285)	8.2	21	14.6	1.76

Table 5.3 presents the mean pause duration, F0-maximum, and articulation rate of the raw prosodic data for each speaker. The table is arranged by gender from the shortest to the longest pause duration.

Table 5.3 Prosodic characteristics per speaker, i.e., for each text (standard deviations in brackets)

	gender of speaker	pause duration (in milliseconds)	F0-maxima (in hertz)	articulation rate (in phonemes per second)
Text 6	female	623 (379)	243 (19)	15.8 (1.5)
Text 3	female	712 (274)	258 (17)	15.2 (1.5)
Text 19	female	724 (244)	324 (24)	13.6 (1.0)
Text 5	female	768 (444)	291 (26)	15.8 (1.0)
Text 14	female	795 (393)	261 (21)	13.9 (1.0)
Text 20	female	894 (405)	308 (36)	14.8 (1.3)
Text 7	female	835 (421)	300 (20)	15.6 (1.0)
Text 11	female	852 (465)	291 (23)	14.1 (1.3)
Text 16	female	907 (319)	268 (21)	14.4 (1.4)
Text 9	female	972 (318)	263 (25)	13.2 (1.4)
Text 2	male	587 (325)	146 (15)	13.7 (1.3)
Text 4	male	743 (171)	161 (12)	15.2 (0.8)
Text 12	male	743 (306)	171 (16)	15.0 (1.3)
Text 15	male	833 (267)	201 (22)	14.8 (1.3)
Text 1	male	827 (418)	153 (20)	14.7 (2.5)
Text 10	male	841 (379)	136 (24)	16.4 (1.3)
Text 18	male	867 (274)	154 (18)	13.5 (1.4)
Text 17	male	919 (208)	193 (19)	12.8 (1.3)
Text 8	male	1273 (476)	186 (23)	13.5 (1.1)
Text 13	male	1496 (437)	188 (27)	15.4 (1.5)

As Table 5.3 shows, there was much individual variation. Therefore, the raw prosodic data were standardized for each speaker. The analyses were performed on these standard scores.

5.3 Results

The central questions are how prosody is related to the three text-structural features, hierarchy, nuclearity, and rhetorical relations. First, however, the effect of the syntactic class of the segments on the prosodic parameters has to be examined, because, as shown in Chapter 4, this is a factor influencing pauses and F0-maxima. For that reason, we first address the effect of syntactic class of the segments on the prosodic characteristics; this is described in section 5.3.1. In sections 5.3.2 to 5.3.4, the investigations of the relations between hierarchy, nuclearity, and rhetorical relations, on the one hand, and the prosodic characteristics, on the other, are reported. Syntactic class is included as a between-groups factor in all analyses.

5.3.1 Effect of syntactic class on prosody

Four syntactic classes were distinguished. First, independent main clauses and main clauses which were the first part of a complex sentence consisting of two coordinate main clauses; these were called ‘main segments in initial position’. Second, main clauses which were the second part of a complex sentence consisting of two coordinate main clauses connected by ‘but’, ‘since’, or ‘and’; these were called ‘coordinate main segments in non-initial position’. Third, subordinate clauses preceding main segments; these were called ‘subordinate segments in initial position’. Fourth, subordinate clauses following main segments; these were called ‘subordinate segments in non-initial position’. In the study described in Chapter 4, subgroups of subordinate segments were not distinguished, because no subordinate segments in initial position occurred in the four texts used in that study. Table 5.4 presents the prosodic means for each of these syntactic classes.

Table 5.4 Pause duration, F0-maximum, and articulation rate (in standard scores) in relation to syntactic class

		preceding pause	F0-maximum	articulation rate
main segments in initial position	(n = 469)	0.22*	0.17	-0.01
coordinate main segments in non-initial position	(n = 47)	-1.15	-0.74	-0.03
subordinate segments in initial position	(n = 10)	0.48	-0.26	-0.24
subordinate segments in non-initial position	(n = 37)	-1.31	-1.09	0.23

*Note: Based on 449 cases since a pause preceding the first segment of each text could not be measured

Separate one-way ANOVA’s for each of the prosodic parameters were run with syntactic class as the independent variable (four levels: main in initial position, coordinate main in non-initial position, subordinate in initial position, subordinate in non-initial position). Syntactic class affected pause duration ($F(3, 539) = 71.77, p<.001, \eta^2 = .29$), and the F0-maximum ($F(3, 559) = 33.72, p<.001, \eta^2 = .15$), but did not affect articulation rate ($F<1$). Pairwise comparisons in

post-hoc analyses (Tukey's HSD procedure) showed that the pauses preceding main segments in initial position and subordinate segments in initial position were significantly longer than the pauses preceding coordinate main segments in non-initial position and subordinate segments in non-initial position. The same pattern was shown for the F0-maximum, except that the F0-maximum of subordinate segments in initial position did not differ significantly from that of main segments in non-initial position.

These results show that it did matter whether a subordinate segment is in initial position or not. Because the number of subordinate segments in initial position was low, and because similarity with Chapter 4 was aimed at, the ten subordinate segments in initial position were not included in the analyses. Main segments in initial position were then regarded as independent segments, and coordinate main segments and subordinate segments in non-initial position as dependent segments. Syntactic class was included in the analyses as an independent factor with two values, 'independent' and 'dependent'.

5.3.2 Effect of hierarchy on prosody

The effect of hierarchy on prosody was evaluated in the same way as in the study described in Chapter 4, i.e., using two variants of the relative approach, and using the absolute approach.

5.3.2.1 Relative approach

5.3.2.1.1 Procedure of linear adjacency

In the relative method of linear adjacency, the prosodic features of all pairs of adjacent boundaries in the texts were examined. In each pair, one boundary was higher in the hierarchical structure than the other. The symmetrical procedure of scoring the levels was applied. Pairs of boundaries with equal levels were excluded from the analyses.

The four texts contained 460 pairs with adjacent boundaries that differed in hierarchical level. Two-way ANOVAs were run to test the effect of hierarchical level on each of the three prosodic parameters. Hierarchical level was included as a within-group factor (two levels: higher, lower) and syntactic combination based on the syntactic classes of the segments in a pair was included as a between-groups factor (four levels: independent-independent, independent-dependent, dependent-independent, dependent-dependent). Separate analyses were performed for the three prosodic parameters: pause duration, F0-maximum, and articulation rate. Table 5.5 presents for each syntactic combination the prosodic means of the higher and lower boundaries in standard scores.

Table 5.5 For each syntactic combination, prosodic characteristics (in standard scores) in relation to hierarchical levels of adjacent boundaries

syntactic class of segments following		hierarchical level of boundary	preceding pause	F0-maximum	articulation rate	n
b_{higher}	b_{lower}					
independent	independent	b_{higher}	0.53	0.24	0.01	315
		b_{lower}	-0.08	-0.04	0.01	
independent	dependent	b_{higher}	0.33	0.03	-0.09	75
		b_{lower}	-1.21	-0.91	0.10	
dependent	independent	b_{higher}	0.51	0.30	0.12	64
		b_{lower}	-1.16	-0.90	0.11	
dependent	dependent	b_{higher}	-1.32	-0.52	0.38	6
		b_{lower}	-1.51	-1.18	0.64	

Note: b_{higher} = highest boundary of the pair; b_{lower} = lowest boundary of the pair

There were effects of hierarchical level ($F(1, 456) = 61.47$, $p < .001$, $\eta^2 = .12$), and syntactic combination ($F(3, 456) = 51.08$, $p < .001$, $\eta^2 = .25$) on the preceding pause. Pauses preceding higher boundaries were longer (mean: 0.47) than those preceding lower boundaries (mean: -0.43). For syntactic combination, all pairwise comparisons (Tukey's HSD procedure) were significant, except that between the independent-dependent and the dependent-independent pair. The more dependent segments involved in a pair, the shorter the corresponding pauses. It did not matter whether the high or the low boundary was followed by a dependent segment. There was an interaction between hierarchy and syntactic class for the preceding pause ($F(3, 456) = 24.93$, $p < .001$, $\eta^2 = .14$). The differences between pauses preceding segments following higher and lower boundaries were different for the four syntactic combinations.

There were effects of hierarchical level on the F0-maximum ($F(1, 456) = 28.39$, $p < .001$, $\eta^2 = .06$). The F0-maxima of segments following higher boundaries were higher (mean: 0.20) than those of segments following lower boundaries (mean: -0.32). There was an effect of syntactic combination on the F0-maximum ($F(3, 456) = 25.17$, $p < .001$, $\eta^2 = .14$). Post hoc analyses showed that, for syntactic combination, all comparisons differed, except those between the independent-dependent combination, on the one hand, and the dependent-dependent combination and the dependent-independent combination, on the other hand. There was an interaction of hierarchy and syntactic class for the F0-maximum ($F(3, 456) = 12.07$, $p < .001$, $\eta^2 = .07$). The F0-maxima patterns of segments following higher and lower boundaries differed for the syntactic combinations.

There were no effects of hierarchical level ($F < 1$) and syntactic combination ($F < 1$) on articulation rate, and no interaction ($F(3, 456) = 1.39$, $p = .25$).

5.3.2.1.2 Procedure of hierarchical adjacency

In the relative method of hierarchical adjacency, the prosodic features of all pairs of dominating boundaries in the hierarchical structure were examined. In each pair, one boundary was higher in the hierarchical structure than the other. Pairs of boundaries with equal levels were excluded from the analyses.

The four texts contained 319 dominating pairs in the analyses of pause duration and 340 dominating pairs in the analyses of the F0-maximum and articulation rate. For the independent-independent combination, fewer pairs were available for the analyses of pause duration, because first segments of the texts were involved in twenty-one dominating pairs. The pauses preceding these segments could not be measured. The same analyses were performed as for the procedure of linear adjacency. Table 5.6 presents for each syntactic combination the prosodic means of the higher and lower boundaries in standard scores.

Table 5.6 For each syntactic combination, prosodic characteristics (in standard scores) in relation to hierarchical levels of dominating boundaries

syntactic class of segments following		hierarchical position of boundary	preceding pause	F0-maximum	articulation rate	n
b _{higher}	b _{lower}					
independent	independent	b _{higher}	0.40	0.32	0.03	260/281
		b _{lower}	-0.25	0.07	-0.04	
independent	dependent	b _{higher}	0.20	-0.07	-0.14	47
		b _{lower}	-1.16	-1.00	0.06	
dependent	independent	b _{higher}	-1.02	-0.38	0.31	7
		b _{lower}	0.16	-0.26	0.63	
dependent	dependent	b _{higher}	-1.16	-0.58	0.40	5
		b _{lower}	-1.15	-0.79	0.40	

Note: b_{higher} = highest boundary of the pair; b_{lower} = lowest boundary of the pair

There was no effect of hierarchical level on the preceding pause ($F < 1$). Syntactic combination affected the preceding pauses ($F(3, 315) = 29.43, p < .001, \eta^2 = .22$). For syntactic combination, all pairwise comparisons (Tukey's HSD procedure) were significant, except that between the independent-dependent combination and the dependent-independent combination. There was an interaction between hierarchy and syntactic combination ($F(3, 315) = 16.89, p < .001, \eta^2 = .14$) for the preceding pause. The differences in pause durations between higher and lower boundaries differed for the four combinations. In the independent-dependent combination and the dependent-independent combination, the differences in preceding pauses between higher and lower boundaries were large, whereas in the independent-independent combination and the dependent-dependent combination, they were small.

There was no effect of hierarchical level on the F0-maximum ($F(1, 336) = 2.79, p = .10$). There was an effect of syntactic combination ($F(3, 336) = 20.96, p < .001, \eta^2 = .16$). For syntactic combination, pairwise comparisons (Tukey's HSD procedure) showed that the F0-maxima differed for the independent-independent combination, on the one hand, and the independent-dependent combination and the dependent-dependent combination, on the other hand. There was an interaction between hierarchy and syntactic combination ($F(3, 336) = 4.31, p < .005, \eta^2 = .04$). The differences in F0-maxima between segments following higher and lower boundaries differed for the four combinations. In the independent-dependent combination, the difference in F0-maxima between higher and lower boundaries was considerable; the differences were minimal in the other three syntactic combinations.

There were no effects of hierarchical level ($F < 1$) and syntactic combination ($F(3, 336) = 1.75, p = .16$) on articulation rate. There was no interaction between hierarchy and syntactic combination for articulation rate ($F < 1$).

5.3.2.2 Absolute approach

In the absolute approach, the prosodic characteristics were evaluated directly in relation to the absolute scores of the levels, defined on an interval scale. Table 5.7 presents for each hierarchical level the mean standard scores of pause duration, F0-maximum, and articulation rate for the independent and the dependent segments. For the independent segments, a one-way analysis of variance was run with hierarchy (five levels) as the independent factor, and each of the three prosodic parameters as the dependent factors. The dependent segments were not analyzed, because of the single observation at level 4, and no observations at levels 3 and 5.

Table 5.7 Prosodic characteristics (in standard scores) of the absolute hierarchical levels of the boundaries per syntactic class (1 = lowest boundary)

syntactic class	level		preceding pause	F0-maximum	articulation rate
independent	level 1	(n = 138)	-0.16	-0.10	-0.05
	level 2	(n = 120)	0.04	0.04	0.11
	level 3	(n = 71)	0.51	0.21	-0.18
	level 4	(n = 74)	0.62	0.28	0.10
	level 5	(n = 45)	0.72	0.39	0.01
dependent	level 1	(n = 72)	-1.23	-0.95	0.08
	level 2	(n = 11)	-1.20	-0.73	0.33
	level 4	(n = 1)	-0.92	1.14	-1.65

For the independent segments, there was an effect of hierarchy on pause duration ($F(4, 443) = 20.34, p < .001, \eta^2 = .16$). The durations of pauses increased as the hierarchical levels increased. The pattern was consistent for independent segments. There was an effect of hierarchy on the F0-maximum ($F(4, 443) = 4.49, p < .001, \eta^2 = .04$). The pattern for the F0-maximum was similar as

that for pause duration: as the hierarchical levels of the boundaries increased, the F0-maxima of the segments following these boundaries increased. The pattern was clear for independent segments. There was no effect of hierarchy on articulation rate ($F(4, 433) = 1.36, p=.35$)

Linear trends were computed on the hierarchical levels for the three prosodic parameters. For the independent segments, the linear trends were significant for pause duration ($F(1, 443) = 76.29, p<.001$) and the F0-maximum ($F(1, 443) = 17.75, p<.001$), but not for articulation rate ($F<1$). For the dependent segments, the linear trend was significant for the F0-maximum ($F(1, 81) = 5.11, p<.05$), but not for pause duration and articulation rate (both $F's<1$).

Correlations were computed for the five hierarchical levels and each of the three prosodic parameters while syntactic class was partialled out. The partial correlation between hierarchy and pause duration was .37 ($p<.01$); between hierarchy and the F0-maximum, .19 ($p<.01$); and between hierarchy and articulation rate, .01 (n.s.). These partial correlations were based on the prosodic realizations of the twenty speakers of the texts. However, the twenty individual speakers differed with regard to the extent to which they realized hierarchy prosodically. Table 5.8 presents the partial correlations for each speaker separately. The table is arranged per gender of the speakers in descending order of the partial correlations between hierarchy and pause durations.

Table 5.8 For each speaker, partial correlations between hierarchy and prosodic characteristics controlled for syntactic class

		pause duration	F0-maximum	articulation rate
female speakers				
speaker 11	(n = 24)	.73 ***	.32	.02
speaker 6	(n = 24)	.60 **	.50 **	-.02
speaker 14	(n = 26)	.57 **	.48 **	-.01
speaker 5	(n = 21)	.57 **	.31	.11
speaker 16	(n = 26)	.46 *	-.27	-.05
speaker 20	(n = 24)	.39 *	-.15	.03
speaker 7	(n = 22)	.24	.11	.39
speaker 19	(n = 29)	.12	.04	-.02
speaker 9	(n = 18)	.06	-.08	.31
speaker 3	(n = 31)	-.07	-.01	.04
male speakers				
speaker 1	(n=23)	.65 ***	.80 ***	-.06
speaker 13	(n = 21)	.63 **	.24	-.17
speaker 12	(n = 28)	.56 **	.01	-.18
speaker 18	(n = 15)	.54 **	.24	-.19
speaker 17	(n = 23)	.45 *	.41 *	.15
speaker 8	(n = 29)	.43 *	.33	-.20
speaker 10	(n = 33)	.38 *	.44 *	.01
speaker 15	(n = 23)	.30	-.06	-.03
speaker 2	(n=22)	.11	-.45 *	.11
speaker 4	(n = 20)	.05	.44 *	.12

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

Partial correlations between hierarchy and the prosodic characteristics differed considerably among the speakers. Thirteen out of the twenty speakers realized longer pauses the higher the boundaries were in the hierarchical structure; the others were not so reliable. Six of the twenty speakers realized a higher F0-maximum in segments which followed higher boundaries; fourteen did not. One of those fourteen speakers, speaker 2, realized a significant opposite pattern for the F0-maximum: he realized a lower F0-maximum in segments which followed higher boundaries. Five speakers realized both pauses and the F0-maximum in the way expected. For none of the speakers did the partial correlation between hierarchy and articulation rate reach significance.

5.3.3 Effect of nuclearity on prosody

This section addresses the question whether prosody was affected by the nuclearity of the segments. First, the distributions of nuclei and satellites are examined. Tables 5.9 and 5.10 present these over the syntactic classes and the hierarchical levels. Table 5.9 contains twenty cases more than Table 5.10, because the first segments of the texts were scored in terms of syntactic class, but the boundaries preceding these first segments could not be scored in terms of hierarchical level. The number of satellites in Table 5.9 and Table 5.10 differs, because the first segment of one text was a satellite.

Table 5.9 Distribution of nuclei and satellites per syntactic class

		independent segments	dependent segments
nucleus	(n=382)	345 (90%)	37 (10%)
satellite	(n=171)	124 (72%)	47 (28%)

Nuclearity and syntactic class were related to each other ($\chi^2(1) = 29.05$, $p < .001$). The proportion of nuclei which were independent segments, was higher than the proportion of satellites which were independent segments: 90 versus 72 percent.

Table 5.10 Distribution of nuclei and satellites per hierarchical level (1 = lowest; 5 = highest)

		level 1	level 2	level 3	level 4	level 5
nucleus	(n=362)	86 (24%)	104 (29%)	64 (17%)	69 (19%)	39 (11%)
satellite	(n=170)	124 (73%)	27(16%)	7 (4%)	6 (4%)	6 (3%)

Nuclearity and hierarchy were related to each other ($\chi^2(4) = 121.56$, $p < .001$). Nuclei occurred more frequently at the higher levels than did satellites.

Table 5.11 presents the prosodic means of nuclei and satellites per syntactic class. Two-way ANOVAs were run with nuclearity (two levels: nucleus, satellite) and syntactic class (two levels: independent, dependent) as independent variables, hierarchy as covariate factor, and the three prosodic parameters, one at a time, as dependent variables. Hierarchy was not included as an independent factor but as a covariate, because of too-low frequencies in some cells of the matrix.

Table 5.11 Prosodic characteristics (in standard scores) of nuclearity per syntactic class

			preceding pause	F0-maximum	articulation rate
independent	nucleus	(n = 325)	0.28	0.13	-0.03
	satellite	(n = 123)	0.05	0.02	0.12
dependent	nucleus	(n = 37)	-1.17	-0.83	-0.17
	satellite	(n = 47)	-1.26	-0.94	0.29

Nuclearity did not affect pause duration ($F < 1$). The pauses preceding nuclei and satellites did not differ in duration. Syntactic class affected pause duration ($F(1, 527) = 126.57, p < .001, \eta^2 = .19$). No interaction was observed (nuclearity \times syntactic class: $F < 1$).

The F0-maxima of nuclei and satellites did not differ ($F < 1$). F0-maxima were higher for independent segments than for dependent segments ($F(1, 527) = 55.41, p < .001, \eta^2 = .10$). There was no interaction (nuclearity \times syntactic class: $F < 1$).

There was an effect of nuclearity on articulation rate ($F(1, 527) = 6.76, p < .01, \eta^2 = .01$). The number of articulated phonemes per second was less for nuclei (mean: -0.06) than for satellites (mean: 0.15): nuclei were read aloud more slowly than satellites. No effect of syntactic class ($F < 1$) was found and there was no interaction ($F(1, 527) = 1.53, p = .21$).

5.3.4 Effects of rhetorical relations on prosody

This section addresses the question whether the prosodic realizations of various rhetorical relations differ. First, it is examined whether causal and non-causal relations have different prosodic characteristics. Second, it is examined whether semantic and pragmatic relations have different prosodic characteristics.

5.3.4.1 Causal and non-causal relations

Tables 5.12 and 5.13 present the distributions of causally and non-causally related segments over the syntactic classes and the hierarchical levels, respectively. The syntactic class of the second segment of the relation is concerned in Table 5.12, because in the operationalization the relation name was assigned to the boundary preceding the second segment of the related pair.

Table 5.12 Distribution of causal and non-causal relations per syntactic class of the second segment of the pair

		second segment independent	second segment dependent
causal relation	(n= 97)	65 (67%)	32 (33%)
non-causal relation	(n=404)	356 (88%)	48 (12%)

There was a dependence between causality and syntactic class ($\chi^2(1) = 25.97, p < .001$). The proportion of independent second segments was lower for causal relations than for non-causal relations: 67 versus 88 percent.

Table 5.13 Distribution of causal and non-causal relations per hierarchical level (1 = lowest; 5 = highest)

		level 1	level 2	level 3	level 4	level 5
causal relation	(n=97)	46 (48%)	27 (28%)	9 (9%)	12 (12%)	3 (3%)
non-causal relation	(n=404)	150 (37%)	96 (24%)	60 (15%)	58 (14%)	40 (10%)

There was no reliable association between causality and hierarchy ($\chi^2(4) = 8.81, p=.07$). Causal and non-causal relations were distributed more or less evenly over the levels.

Table 5.14 presents the prosodic characteristics of causal and non-causal relations per syntactic class of the second segment. Two-way ANOVAs were run with causality (two levels: causal, non-causal) and syntactic class (two levels: independent, dependent) as independent variables, hierarchy as a covariate factor (five levels), and each of the three prosodic parameters as dependent variables. Hierarchy was not included as an independent factor but as a covariate, because of too-low frequencies in some cells of the matrix.

Table 5.14 Prosodic characteristics (in standard scores) of causally and non-causally related segments per syntactic class of the second segment of the pair

			preceding pause	F0-maximum	articulation rate
independent second segment	causal	(n = 65)	-0.01	-0.04	0.16
	non-causal	(n = 355)	0.28	0.12	-0.03
dependent second segment	causal	(n = 32)	-1.32	-0.76	0.32
	non-causal	(n = 48)	-1.15	-1.00	-0.06

Pauses between causally related segments were shorter than those between non-causally related segments (-0.44 versus -0.11; $F(1, 495) = 4.54, p<.05, \eta^2 = .01$). Pauses were longer preceding independent segments than preceding dependent segments ($F(1, 495) = 98.48, p<.001, \eta^2 = .17$). There was no interaction ($F<1$).

Causality did not affect the F0-maximum ($F<1$). The syntactic class of the second segments affected the F0-maxima ($F(1, 496) = 40.01, p<.001, \eta^2 = .08$). There was no interaction ($F(1, 496) = 2.43, p=.12$).

Articulation rate was affected by causality ($F(1, 495) = 4.99, p<.05, \eta^2 = .01$). Causally related segments were read faster than non-causally related segments: the mean articulation rate of causally related segments was 0.22; the mean articulation rate of non-causally related segments was -0.03. There was no effect of syntactic class and no interaction (both F 's <1).

5.3.4.2 Semantic and pragmatic relations

Tables 5.15 and 5.16 present the distributions of semantically and pragmatically related segments over the syntactic classes and the hierarchical levels, respectively. The total number of semantic and pragmatic relations was lower than the total number of causal and non-causal relations, because the Joint relation was classified neither as semantic nor as pragmatic and excluded from these analyses.

Table 5.15 Distribution of semantic and pragmatic relations per syntactic class of the second segment of the pair

		second segment independent	second segment dependent
semantic relation	(n= 305)	266 (87%)	39 (13%)
pragmatic relation	(n=88)	68 (77%)	20 (23%)

There was a dependence between semantic and pragmatic relations, and the syntactic class of the second segment of these relations ($\chi^2(1) = 5.29$, $p < .05$). The proportion of independent second segments was higher for semantic relations than for pragmatic relations: 87 versus 77 percent.

Table 5.16 Distribution of semantic and pragmatic relations over the hierarchical levels (1 = lowest; 5 = highest)

		level 1	level 2	level 3	level 4	level 5
semantic relation	(n= 305)	117 (38%)	75 (25%)	43 (14%)	42 (14%)	28 (9%)
pragmatic relation	(n=88)	30 (34%)	21 (24%)	12(14%)	14 (16%)	11 (12%)

There was no association between semantic and pragmatic relations, and hierarchy ($\chi^2(4) = 1.34$, $p = .86$).

Table 5.17 presents the prosodic characteristics of the semantic and pragmatic relations per syntactic class of the second segment. Two-way ANOVAs were run with semanticity (two levels: semantic, pragmatic) and syntactic class (two levels: independent, dependent) as independent variables, hierarchy as a covariate, and each of the three prosodic parameters as dependent variables. Hierarchy was not included as an independent factor, because of too-low frequencies in some cells of the matrix.

Table 5.17 Prosodic characteristics (in standard scores) of semantic and pragmatic relations per syntactic class of the second segment of the pair

			preceding pause	F0-maximum	articulation rate
independent second segment	semantic	(n = 266)	0.16	0.01	0.06
	pragmatic	(n = 68)	0.28	0.24	0.02
dependent second segment	semantic	(n = 39)	-1.33	-0.92	0.23
	pragmatic	(n = 20)	-1.01	-0.74	0.31

Semanticity did not affect pause duration ($F(1, 388) = 2.47$, $p = .12$). Pause duration did not differ for semantically and pragmatically related segments. Syntactic class did affect pause duration ($F(1, 388) = 65.57$, $p < .001$, $\eta^2 = .15$). Independent second segments had longer preceding pauses than dependent second segments. There was no interaction between semanticity and syntactic class ($F(1, 388) = 2.06$, $p = .15$).

Semantically related segments did not have different F0-maxima than pragmatically related segments ($F(1, 388) = 1.82, p=.18$). Syntactic class did affect the F0-maximum ($F(1, 388)=25.59, p<.001, \eta^2=.06$). Independent second segments had higher F0-maxima than dependent second segments. There was no interaction between semanticity and syntactic class ($F<1$).

Articulation rate was not affected by semanticity ($F<1$), nor by syntactic class ($F(1, 388)=1.87, p=.17$). There was no interaction (semanticity x syntactic class: $F<1$).

5.4 Conclusion and discussion

The aspects of text structure captured using Rhetorical Structure Theory, i.e., hierarchy, nuclearity, and rhetorical relations, were found to be related to prosody in various ways. Table 5.18 presents an overview of these results.

Table 5.18 Overview of the effects of hierarchy, nuclearity, and rhetorical relations on pause duration, F0-maximum, and articulation rate

		preceding pause	F0-maximum	articulation rate
Hierarchy				
Relative approach	linear adjacency	$p<.001$	$p<.001$	-
	hierarchical adjacency	-	-	-
Absolute approach		$p<.001$	$p<.001$	-
Linear trend	independent segments	$p<.001$	$p<.001$	-
	dependent segments	-	$p<.05$	-
Partial correlations		$p<.01$	$p<.01$	-
Nuclearity		-	-	$p<.01$
Rhetorical relations				
Causality		$p<.05$	-	$p<.05$
Semanticity		-	-	-

The relative procedure of linear adjacency, the absolute approach, and the linear trend for independent segments showed that the speakers indicated the hierarchical structure of a text using pause duration and pitch range: higher boundaries in the text structure have longer pauses than lower ones, and the segments following higher boundaries have higher F0-maxima than lower ones. In contrast, the procedure of hierarchical adjacency did not show main effects of hierarchy on pause duration, and on the F0-maximum. This may partly be explained by the smaller number of observations in the procedure of hierarchical adjacency. In the procedure of linear adjacency, dependent segments dominated independent segments 64 times since all adjacent pairs of segments were involved in the analyses, whereas in the procedure of hierarchical adjacency, dependent segments dominated independent segments only 7 times. For this syntactic combination the high-low pattern was reversed for the two procedures. The different results may also be explained by the procedures themselves: the results of the procedure of linear adjacency

may be considered reflections of the incremental processing of the text to a higher extent than the results of the procedure of hierarchical adjacency.

The results with regard to the relation between hierarchy and pause duration were in accordance with Schilperoord (1996). He showed that, in dictated speech, pause durations were shorter when boundaries were more subtle. The hierarchical levels defined in this study were probably strongly related to the various boundary types defined by Schilperoord, for instance, 'main transitions' are comparable with high boundaries and 'incidental transitions' are comparable with low boundaries. The effect of hierarchy on the F0-maximum found in this study extends Schilperoord's findings with regard to pause duration. By systematically varying pause duration and pitch range, speakers show that in a way they are aware of the various levels of a text structure. By this variation in pausing and pitch range, they may facilitate the listener's task of processing the segmentation and the hierarchical levels of a text. Assuming that speakers intend to facilitate the listener's decoding task, the prosodic characteristics of the various levels of the text structure are signals that enable the listener to more easily distinguish main transitions from incidental transitions. It should be noted, however, that explanations in terms of a listener's perspective are speculative as long as it has not been investigated whether listeners really perceive the prosodic differences between the hierarchical levels. Further research must be concerned with listener's perceptions of the prosodic marking of text-structural aspects.

Fragments of the sample text presented in Table 5.1 may explain the general pattern of gradually decreasing pause durations and F0-maxima from higher to lower levels. In segment 7, the writer notes the difficulty of finding people at home. This point is illustrated using four groups of people: farmers (8-12), married people (13-16), homeless people (17-21), and people who argue that their privacy is affected (22-24). The text analysis shows that the four text parts are four separate arguments for segment 7, but that they also cohere as one argument. In the sequence of segments 7- 25, the transitions between 7-8, 12-13, 16-17, 21-22, and 24-25 had longer pause durations and a higher F0-maximum than did the transitions within each text part. The transition between 24 and 25 was prosodically realized even more strongly, since a new topic is introduced there, namely, the solution for the problem posed, and, therefore, a higher level in the hierarchy is reached. Another example of a clear high-low pattern is boundary 5-6. At that point, the writer starts to retell more extensively what was summarized in the first five segments of the lead of the article, and, therefore, there is a change upwards in hierarchical level. There were longer pause durations and higher F0-maxima in the first segment following this boundary.

Individual speakers differed greatly in the extent to which they marked the hierarchical level of the text structure with pauses and F0-maxima. Hierarchy was correlated to pause duration for thirteen speakers; and to pitch for seven speakers, one of whom showed an inversed correlation. There was no correlation with any of the prosodic parameters for six speakers. For five speakers was hierarchy related to both pause duration and the F0-maximum. One of the reasons for the individual variation in prosodically realizing text structure may be that 'good' and 'bad' speakers were not distinguished. With regard to the goal of this study, i.e., demonstrating the relation between text structure and prosody, 'bad' speakers would have been those speakers who read aloud monotonously, and speakers who were insensitive to all text-structural notions. If there

were 'bad' speakers among the twenty speakers, they diminished the effect of hierarchy on prosody. An example of a 'bad' speaker was Speaker 2. He differed from the other speakers in that his F0-maximum increased instead of decreased from the highest level to the lowest level of text structure. In general, if bad speakers had been removed from the analyses the relation between text structure and prosody may have been demonstrated far more strongly (see also Noordman, Dassen, Terken, & Swerts, 1999). Distinguishing 'good' and 'bad' speakers should have to be done on grounds independently from prosody.

Other explanations for the individual differences are that speakers use different prosodic characteristics to indicate text structure from those measured in this study, for example, loudness or vowel lengthening, or that speakers have difficulties in reading texts aloud. In order to examine more clearly the individual differences between speakers in their prosodic realizations of the hierarchical structure of the text, they should have been asked to read aloud the same text.

The speakers indicated nuclearity using articulation rate. Nuclei were read at a slower rate than satellites. Satellites are less important for the coherence of the text. Even if they are left out, the content of a text can be understood. Speakers might 'know' that fast readings of the satellites do not prevent the listener from a clear understanding, provided that the variation in articulation rate is perceptually relevant. Listeners may be helped by slow readings of nuclei to interpret this information as more important.

The speakers indicated causality using both time-related characteristics: pause duration and articulation rate. In the sample analysis presented in Figure 5.1, the relations between segments 2 and 3, segments 10 and 11, segments 11 and 12, segments 13 and 14, and segments 24 and 25 were characterized as causal ones. The pauses between those segments were shorter than the pauses between the other segments; and the second segments of those pairs were read faster than the other segments. These results may be interpreted in accordance with the findings of Sanders and Noordman (2000). They showed that causal relations need a shorter processing time than non-causal relations, although causal relations are more complex and more informative. Our results show that a shorter processing time seems also to be manifested in the production of speech. The shortening of pauses and increase of articulation rate may indicate that causally related segments cohere more strongly than non-causally segments.

A critical remark should be made here. The findings of the statistical analyses may suggest that the causal relations in the texts occurred between two adjacent segments in the texts. However, the text structure in Figure 5.1 shows that the causal relations occurred not simply between adjacent segments, but rather between larger text spans. For example, the text span containing segments 11 and 12 was causally related to the text span containing segments 8 to 10. As a result of the way the rhetorical relations were operationalized, the causal relation was associated with the boundary between segments 10 and 11, because this was the boundary between both causally related text spans. Strictly speaking, there was no causal relation between segments 10 and 11 at all; whereas, nevertheless, the prosodic characteristics of this boundary and the segment following the boundary were measured. In experimental follow-up research on the prosodic realization of causality, causal and non-causal relations must be constructed between

two segments and not between larger text spans, so that the boundary exactly corresponds with the connection between the two related segments.

The semanticity of segments did not affect any of the three prosodic parameters. The same critical remarks can be made concerning the semantic and pragmatic relations as were made regarding the causal and non-causal relations. The semantic and pragmatic relations in the text material occurred not simply between adjacent segments, but rather between larger text spans. For example, the Motivation relation between segments 6 and 7 was not restricted to these two segments, but concerned segment 6, on the one hand, and the text span consisting of segments 7 to 24, on the other hand. Another point is that the number of pragmatically related segments in the text material, about one quarter of the total number of relations, appeared to be high in the newspaper reports which were intended to describe events in an objective way. Although we used Mann and Thompson's list of 'subject matter' and 'presentational' relations as a criterion for strictly classifying the relations as semantic and pragmatic, the criterion may have had a bias for pragmatic relations. To enable a pure demonstration of the effect of semanticity on prosody, follow-up research on the prosodic realization of semantic and pragmatic relations should be concerned with a systematic manipulation of this relation type.

The various aspects of text structure affect prosody significantly. Of the three text-structural features investigated, hierarchy, nuclearity, and rhetorical relations, hierarchy was marked most strongly by prosody. Effects of nuclearity and causality were found, but the explained variances of these factors were very low. An explanation might be that the scope to vary prosody was too small to signal all text-structural information. For example, when speakers lengthen their pauses to indicate syntactic and hierarchical information, it is impossible to lengthen the pauses even more to indicate a particular rhetorical relation as well. This would also be communicatively unacceptable to listeners, whether or not they were able to perceive the prosodic differences between the various aspects of text structure: When listeners perceive a short pause, how do they know whether it marks a low hierarchical level or a causal relation or both? Further research must address such questions.

The syntactic class of the segments was controlled for in all statistical analyses. Syntactic class was found to be an important factor to take into account in the study of the relation between text structure and prosody. In addition, the analysis described in section 5.3.1 showed that initiality also is a factor of interest. In the present study, subordinate segments in initial position were removed from the statistical analyses, because their prosodic realizations looked more like main segments in initial position than like subordinate segments in non-initial position and, therefore, they would have been confounding with the effect of syntactic class on the prosodic parameters. In further research, the position of segments must also be taken into consideration.

The main questions raised in this study concerned how the various hierarchical levels of the boundaries, the nuclearity of the segments, and the particular rhetorical relations between segments are reflected in prosody. Although it is difficult to capture the results for the various

text-structural aspects in a particular theoretical model which coherently explains why the conceptual structure of a text is realized prosodically in these ways, the results at least provide evidence that the prosodic realization of text structure is more subtle than has been demonstrated by research in which the structure of a text was considered merely a succession of sentences and paragraphs which are realized in prosodically different ways. The findings of this study show that texts have to be considered multilayered structures, consisting of more and less important elements which are related to each other by means of particular rhetorical relations; and that pause duration, and the F0-maximum, are strong means for speakers, but not for all speakers, to express these structural characteristics of texts.

6

Prosody of causal and non-causal, and of semantic and pragmatic relations: Two experiments

6.1 Introduction

The focus of this study is on the prosodic realization of causal and non-causal relations and semantic and pragmatic relations. To investigate whether there were prosodic differences between causal and non-causal relations, and between semantic and pragmatic relations, in the study described in Chapter 5, all rhetorical relations in the text corpus were categorized as either causally or non-causally related and as either semantically or pragmatically related. There were several confounding factors inherent in this approach. First, segments differed with regard to their content and length, and they occurred at different positions in the linear order of the texts and in the hierarchy of the text structure. Second, rhetorical relations did not always occur between two adjacent segments, but often between two larger text parts or between a larger text part and a single segment. Third, causally and non-causally related segments, and semantically and pragmatically related segments, occurred both marked and unmarked by connectives. Fourth, rhetorical relations occurred in different linear orders: a causal relation could occur in both cause-result and result-cause order, and a pragmatic relation could occur in both fact-conclusion and in conclusion-fact order. Each of these four factors may have had an influence on prosodic marking.

In order to get a clear view of the relation between rhetorical relations and prosody, two experiments are run using constructed text materials. In one experiment, the texts contain target sentences which are either causally or non-causally related to their preceding sentences; in the other experiment, the texts contain target sentences which are either semantically or pragmatically related to their preceding sentences. The target sentences are identical in the two conditions. The texts are read aloud by speakers who have carefully prepared the reading task. In each experiment, the prosodic characteristics of the target sentences in both conditions are measured. The experiment on causal and non-causal relations is reported in section 6.2; the experiment on semantic and pragmatic relations in section 6.3. Both experiments include a pretest to investigate the adequacy of the text materials and a main study to examine the prosody of the types of rhetorical relations under investigation.

6.2 Experiment 1: Prosody of causal and non-causal relations

In the research reported in Chapter 5, all rhetorical relations in the twenty texts were categorized as either causally or non-causally related using Mann and Thompson's criteria (1988). These categories can also be described in terms of the 'basic operation' of a coherence relation, i.e., a segment is related to another segment either causally or non-causally (Sanders, Spooren, & Noordman, 1992). The results reported in Chapter 5 show that causally related segments are preceded by shorter pauses and read aloud at a higher articulation rate than are non-causally related segments. The factors discussed in the preceding section, however, make the conclusions about the effect of the basic operation preliminary. In the present study, the effect of type of rhetorical relation is assessed under experimental control.

6.2.1 Pretest: Construction and selection of text material

There were twenty-seven target sentences. Each target sentence was included in two texts. In one of the texts, the target sentence was causally related to its preceding sentence; in the other text, it was non-causally related to its preceding sentence. Twenty-seven pairs of texts resulted. A pretest of the text material was conducted to test whether text manipulation succeeded in creating causal and non-causal interpretations of the target sentences.

6.2.1.1 Text material

The texts were derived from news reports in newspapers and weekly magazines. Complex linguistic constructions and unfamiliar words were avoided. In each pair of texts, the same target sentence was included, which was either causally or non-causally connected with its preceding sentence. For a good comparison of their prosodic characteristics, the target sentences had to be identical in the two conditions. Therefore, connectives were avoided (such as *therefore*, *as a result of* in the causal condition and *and*, *thirdly* in the non-causal condition). Each text consisted of six or seven sentences: the target sentence was preceded by three or four sentences, and followed by at least one sentence. The causal and non-causal conditions of a text are illustrated in (1) and (2). In the examples, the target sentences are printed in bold; the preceding sentence to which a target sentence is related is printed in italics.

- (1)
causal relation

Necessary investments have caused Dutch Railways to be in the red. *In a press conference they admitted that many materials have to be replaced.* **Trains and busses will become considerably more expensive.** Reactions of consumer organizations to this price increase are not yet known. The measure is very inconvenient because the use of public transport had just started to be stimulated by various agencies.

[Noodgedwongen investeringen kleuren de cijfers van NS rood. *In een persconferentie gaven ze toe dat veel van het materieel vervangen moet worden.* **Trein en bus worden fors duurder.** Het is nog niet bekend hoe consumentenorganisaties reageren op deze actie. De maatregel komt erg slecht uit omdat door verschillende instanties het gebruik van openbaar vervoer gestimuleerd zou gaan worden.]

- (2)
non-causal relation

Inflation is felt in the entire tourist sector, in both transport and accommodation. Hotels almost double their prices. *Air companies drive their prices up substantially.* **Trains and busses will become considerably more expensive.** The prices of apartments are no longer comparable with the prices of a few years ago. These are only a few of the complaints which have reached the ANVR during the last three months.

[De inflatie is voelbaar in de gehele toeristische sector, zowel in vervoer als verblijf. De hotels verdubbelen bijna de prijzen. *Vliegmaatschappijen schroeven de prijzen behoorlijk op.* **Trein en bus worden fors duurder.** Appartementen zijn qua prijs niet meer vergelijkbaar met een aantal jaren geleden. Dit is nog maar een greep uit de grote berg aan klachten, die de ANVR de afgelopen drie maanden heeft ontvangen.]

The target sentence can be paraphrased in (1) as ‘Therefore, trains and busses will become considerably more expensive’; in (2) as ‘And (or: ‘Thirdly,...’ or ‘Also,...’) trains and busses are getting considerably more expensive’. Causal relations may be more difficult to recognize, because the relation has to be inferred owing the lack of connectives. For example, in (1) the causality of the relation has to be inferred in the following way: ‘To be able to bear the costs of the replacements, trains and busses will become considerably more expensive’.

Other factors were also held constant in the construction of target sentences. For all causally related target sentences, the direction of causality was held constant, i.e., the target sentence always contained the result or the solution, and its preceding sentence always contained the cause or the problem. For all non-causally related target sentences, the form of the addition was held constant, i.e., the target sentence was always the third element in a list of four items. The number of new topics and the hierarchical level of the target sentence within the text structure was held constant for the two conditions.

6.2.1.2 Judges

Forty-one persons participated in the pretest. They were native speakers of Dutch. Because broad reading experience and text understanding were required, people with at least Higher Vocational Education were selected. Two thirds of the participants were students and teachers of the Dutch School of Tourism, Department of Communication, in Breda; one third were students from the Faculty of Arts at the Vrije Universiteit Amsterdam. Their ages ranged from 18 to 59 years; the mean age was 33.5 years. They were not paid for the task.

6.2.1.3 Procedure

The pretest consisted of the twenty-seven pairs of texts in a causal and a non-causal condition, the causality and plausibility of which had to be judged. The pairs were distributed over lists such that the causal condition of a text pair did not co-occur with the non-causal condition. The twenty-seven pairs of text were distributed over four lists consisting of six, seven, or eight pairs of texts. Each judge received one of these lists.

In the causality test, the judges indicated whether the relation between the target sentence and its preceding sentence was causal or non-causal on a dichotomous scale. The distinction between causally and non-causally related sentences was explained in the instruction using examples. These instructions are presented in Appendix D. In the plausibility test, judges indicated to what extent the target sentence followed its preceding sentence plausibly on a five-point scale ranging from ‘very unnaturally’ (1) to ‘very naturally’ (5). These instructions are presented in Appendix E. In the lists, the target sentences were printed in bold to indicate that judges had to examine the rhetorical relation between that sentence and its preceding sentence. The judges were not allowed to communicate with each other about the task. The task was self-paced.

6.2.1.4 Results

Texts to be used in the main study were selected on the basis of the results of both the causality test and of the plausibility test. The causality test was considered more important, because the causal or non-causal relation between a target sentence and its preceding sentence was the independent variable in the main study. The criterion was that at least 80 percent of the judges should agree regarding the causality or non-causality of the rhetorical relation. Out of the twenty-seven texts, six causal texts and three non-causal texts did not reach this agreement level. One of the texts scored below 80 percent in both conditions. Therefore, eight pairs of texts were removed.

The criterion with regard to plausibility scores was that the conditions should not differ significantly. Out of the remaining nineteen pairs of texts, five pairs differed with regard to plausibility in the two conditions. Because at least sixteen pairs of texts were required in the main study, the three with the lowest plausibility scores were removed. The other two texts were rewritten. From the remaining sixteen pairs of texts, causally related target sentences scored a mean of 2.56 ($sd = .56$), with a range from 1.56 to 3.50, in the plausibility test, and non-causally related target sentences 2.78 ($sd = .57$), with a range from 1.33 to 3.50. Conditions did not differ with regard to plausibility ($t(15) = 1.64, p = .12$). The selected texts were considered clear manipulations of causal and non-causal relations.

6.2.2 Main study: Prosodic realization of causal and non-causal relations

6.2.2.1 Speakers

Twenty-five speakers participated in the experiment: twenty-three students of the faculties of arts, law, and social sciences from Tilburg University, and two teachers of the Dutch School of Tourism, Department of Communication, in Breda. They were native speakers of Dutch. There were seven male and eighteen female speakers. Their ages ranged from 19 to 38 years; the mean age was 23.5 years. The speakers were not informed about the goal of the experiment, and they were not paid for their participation.

6.2.2.2 Procedure

The sixteen pairs of texts are presented in Appendix G. Each text was printed on a separate sheet using a 14-point font. The target sentences and their preceding sentences were printed in the same font as the surrounding text. The texts were presented without paragraph markers. The speakers were encouraged to prepare the reading task thoroughly and to make notes in the texts, for example, to mark paragraph breaks or certain words. The speakers were instructed to read the texts aloud as if they were newsreaders at a radio station, and to read them aloud without speech errors. There was one training text. The speakers read aloud both versions of each text to make the comparability of the conditions maximal. The experiment was run in two blocks of sixteen texts. The order of the texts was arranged so that, for each pair, one condition was placed before the break, and the other after it. To avoid an influence of order, two lists with reversed orders

were used. In the break between the blocks, some personal data of the speakers were registered and the speakers were asked to answer a questionnaire about their reading capacities. The recordings took place in a sound-proof room. Speech was recorded directly using a portable personal computer, and digitized using the speech-processing program Gipos. Recordings were started again from the beginning of a text when speech was not fluent. This happened on average twice for each speaker. A reading session lasted about twenty minutes.

6.2.2.3 Speech material

The prosodic characteristics were the same as those in the study reported in Chapter 5: duration of the preceding pause, the F₀-maximum, and articulation rate of the target sentence. In addition, pause duration following the target sentence was measured, because it was thought that pauses preceding and following a target sentence may be related in some way. In addition to the F₀-maximum, the mean pitch range of the whole target sentence was also measured, because it was thought that the whole pitch contour of the target sentence might be different in the two conditions. A different contour may result in a different mean pitch range, rather than a different F₀-maximum.

6.2.2.4 Results

Table 6.1 presents pause durations preceding and following the target sentence, and the F₀-maximum, mean pitch range, and articulation rate of the target sentence for non-causal and causal relations.

Table 6.1 Prosodic characteristics of causal and non-causal relations

	causal	non-causal	F ₁	F ₂
pause preceding (in milliseconds)	592	555	***	
pause following (in milliseconds)	518	544		
F ₀ -maximum (in hertz)	247.7	245.3	*	
mean pitch range (in hertz)	182.1	179.7	***	**
articulation rate (in phonemes per second)	15.8	15.7		

Note: * p<.05; ** p<.01; *** p<.001

F₁ and F₂ analyses of variance with repeated measures were run. The independent factor was type of rhetorical relation (two levels: causal, non-causal). Speaker was the random variable in the F₁ analysis, target sentence in the F₂ analysis. The order of the texts was included as a between-groups factor (two levels: text 1-32, text 32-1). Because order was not a factor of interest in any of the analyses, no results of order are reported.

The pause preceding a target sentence was on average 37 milliseconds longer between causally related sentences than between non-causally related ones. In the F₁ analysis of variance,

the effect on pause duration preceding the target sentence was significant; in the F_2 analysis, it was not ($F_1(1, 24) = 19.15, p < .001, \eta^2 = .44$; $F_2(1, 15) = 1.88, p = .19$).

The pauses preceding and following the target sentences in the two conditions were compared. Therefore, an additional analysis of variance was run with two within-group factors: Location of pause (two levels: preceding, following) and Rhetorical relation (two levels: causal, non-causal). There was a main effect of Location ($F_1(1, 24) = 39.94, p < .001, \eta^2 = .63$; $F_2(1, 15) = 5.72, p < .05, \eta^2 = .28$). There was an interaction between type of rhetorical relation and location of the pauses ($F(1, 24) = 13.18, p < .01, \eta^2 = .35$; $F_2(1, 15) = 4.03, p = .06, \eta^2 = .21$): in causal relations, pauses preceding target sentences were longer than those following target sentences ($F_1(1, 24) = 33.48, p < .001, \eta^2 = .58$; $F_2(1, 15) = 14.17, p < .01, \eta^2 = .49$), whereas in non-causal relations, pauses preceding and following target sentences did not differ ($F_1(1, 24) = 1.36, p = .26$; $F_2 < 1$). The pause following the target sentence did not differ for causal and non-causal relations ($F_1(1, 24) = 3.89, p = .06, \eta^2 = .20$; $F_2(1, 15) = 1.46, p = .25$).

The F0-maximum of a target sentence which was causally related to its preceding sentence was on average 2 hertz higher than that of a target sentence which was non-causally related to its preceding sentence ($F_1(1, 24) = 6.97, p < .05, \eta^2 = .23$; $F_2(1, 15) = 1.28, p = .28$). The mean pitch range of causally related target sentences was on average also 2 hertz higher than that of non-causally related target sentences ($F_1(1, 24) = 16.86, p < .001, \eta^2 = .41$; $F_2(1, 15) = 7.77, p < .05, \eta^2 = .34$).

The type of rhetorical relation did not affect the articulation rate of the target sentence ($F_1 < 1$; $F_2 < 1$).

Table 6.2 presents for each speaker the difference scores of the prosodic measurements between causal and non-causal relations. The scores of causal relations were subtracted from the corresponding scores of non-causal relations. Articulation rate is not included in the table, because it was not affected by type of rhetorical relation. In addition to the analyses of variance, the difference between causal and non-causal relations was tested using Wilcoxon-matched-pairs tests.

Table 6.2 For each speaker, difference between non-causal and causal relations in pause duration (in milliseconds), F0-maximum, and mean pitch range (in hertz)

	preceding pause	following pause	F0-maximum	mean pitch range
speaker 1	- 72	- 9	+ 2.4	+ 1.8
speaker 2	- 37	- 60	- 0.1	- 0.7
speaker 3	- 2	- 17	+ 5.0	- 1.9
speaker 4	- 87	+ 57	+ 5.5	- 0.3
speaker 5	- 34	- 62	+ 1.3	- 1.6
speaker 6	- 29	+ 100	- 1.8	- 4.2
speaker 7	- 48	+ 28	- 4.8	- 4.3
speaker 8	- 10	+ 14	- 1.8	- 1.2
speaker 9	- 15	+ 79	- 1.0	+ 0.1
speaker 10	- 20	- 22	- 10.2	- 6.5
speaker 11	- 57	- 21	- 1.9	- 4.7
speaker 12	- 19	- 12	+ 0.8	- 0.6
speaker 13	- 150	+ 219	- 0.7	+ 1.2
speaker 14	- 54	+ 57	- 5.5	- 2.3
speaker 15	- 122	+ 89	- 7.9	- 9.9
speaker 16	- 70	- 35	- 8.4	- 6.2
speaker 17	+ 13	+ 82	+ 2.3	- 1.6
speaker 18	- 23	+ 21	- 9.8	- 5.0
speaker 19	- 8	- 23	- 3.4	- 1.6
speaker 20	+ 30	- 11	+ 3.2	+ 0.2
speaker 21	- 56	+ 28	- 7.1	- 2.8
speaker 22	- 1	+ 95	- 8.9	- 6.9
speaker 23	+ 36	+ 106	+ 1.6	- 2.3
speaker 24	- 64	- 18	- 2.6	+ 2.0
speaker 25	- 38	- 37	- 0.9	- 1.0
non-causal < causal	N = 22	N = 12	N = 17	N = 20
non-causal > causal	N = 3	N = 13	N = 8	N = 5

Twenty-two speakers produced longer pauses preceding causally related target sentences than preceding non-causally related ones ($z = 3.57$, $p < .001$). Pause duration following the target sentence did not differ significantly for the two conditions ($z = 1.53$, $p = .13$). The vast majority of speakers read aloud causally related target sentences with a higher F0-maximum ($z = 2.33$, $p < .05$) and a higher mean pitch range ($z = 3.38$, $p < .001$).

Table 6.3 presents for each text the difference scores of the prosodic measurements. Only the effect on mean pitch range was found to be significant ($z = 2.48$, $p < .05$).

Table 6.3 For each text difference between non-causal and causal relations in pause duration (in milliseconds), F0-maximum, and mean pitch range (in hertz)

	preceding pause	following pause	F0-maximum	mean pitch range
text 1	+ 2	+ 190	- 3.9	- 4.0
text 2	- 130	- 17	- 14.6	- 6.7
text 3	+ 33	+ 55	- 0.4	- 0
text 4	+ 140	- 66	+ 3.1	+ 2.2
text 5	+ 14	- 30	- 7.8	- 5.9
text 6	- 38	+ 26	- 1.0	- 3.0
text 7	- 29	- 4	- 4.6	- 3.6
text 8	+ 54	+ 150	- 3.0	- 1.4
text 9	+ 46	+ 170	+ 9.6	- 0.7
text 10	- 92	+ 31	- 9.5	- 6.4
text 11	- 261	- 47	- 19.2	- 9.5
text 12	- 83	- 29	- 2.1	+ 3.3
text 13	- 215	+ 20	+ 5.1	- 1.3
text 14	- 127	- 11	- 7.8	- 0.5
text 15	+ 88	- 111	+ 6.0	- 1.1
text 16	+ 5	+ 87	+ 11.9	+ 0.1
non-causal < causal	N = 8	N = 8	N = 11	N = 13
non-causal > causal	N = 8	N = 8	N = 5	N = 3

6.2.2.5 Discussion and conclusion

Mean pitch range was found to be affected by the causality of the rhetorical relation: the mean pitch range of causally related sentences was higher than that of non-causally related sentences. Two hertz may be a small difference and, from a perception perspective, insignificant. From a production perspective, however, the effect was found to be consistent: twenty out of the twenty-five speakers produced causally related sentences with a higher mean pitch range. The pauses between causally related sentences were found to be longer than those between non-causally related sentences, and these sentences also had a higher F0-maximum. The results with regard to the preceding pause and the F0-maximum may be generalized over speakers, but not over target sentences. The pause pattern in the non-causal condition suggests that the speakers read aloud the items in a rhythmical way: the durations of the pauses preceding the third and fourth parts of the enumeration were constant. The pause pattern in the causal condition shows that the speakers stopped speaking for a short while between the two sentences, the first of which was intended to present the cause or problem, and the second of which was intended to present the result or solution.

These results differed from the results of the corpus study reported in Chapter 5. Those results showed that pauses were shorter and the articulation rate was faster for causally related sentences than for non-causally related sentences. Pitch range was not affected by causality. One explanation for the conflicting results with regard to pause duration may be that prosody was

realized differently in sentences whose rhetorical relations are lexically marked, as in the texts used in the corpus study, than in sentences whose rhetorical relations are not lexically marked, as in the text used in the experiment. Because of the lack of lexical markers, i.e., either connectives or content cues, it might be assumed that another marker of the causal relation was needed, i.e., a longer pause duration. This would be a valid explanation if the plausibility of the causal relations was lower than that of non-causal relations. That was not the case, however, at least not for the final selected set of texts. In the original set of twenty-seven constructed pairs of texts, the plausibility of the causal relations was lower than that of the non-causal relations; for the final selection, it was necessary that the causal relations and the non-causal ones be equally plausible.

Table 6.4 shows that there is a relation between plausibility and causality relations. It presents for the causal and non-causal condition the correlations between plausibility, on the one hand, and preceding and following pauses, the F0-maximum, mean pitch range, and articulation rate, on the other hand.

Table 6.4 For causally and non-causally related target sentences correlations between plausibility and prosodic characteristics

		preceding pause	following pause	F0-maximum	mean pitch range	rate
non-causal	(n=16)	.02	.01	-.03	-.15	.12
causal	(n=16)	-.47*	-.17	.06	-.04	-.19

* Correlation is significant at the .05 level (one-tailed)

For the texts in the causal condition a significant correlation between plausibility and preceding pause was found. The durations of the pauses between causally related sentences increased as the plausibility of the sentences decreased. This correlation was not found for non-causally related sentences. If we assume that the readers did not recognize the causal relations in the causal texts with low plausibility, the lengthening of the pause preceding causally related sentences may also be explained as an effect of hierarchy: in these cases, the speakers may have realized a higher-level boundary in the text structure resulting in a longer preceding pause. For causal texts with high plausibility, the negative correlation indicates that pause durations got shorter. Inspection of the data suggests that, for plausibility ranges beyond 3.5 (the upper limit), the pause durations for causal texts were in fact shorter than for non-causal items, which is consistent with the findings in Chapter 5. On the basis of this reasoning, we expect that the presence of lexical markers and content-like cues would have a major influence on the pause patterns for causal relations. This is clearly speculative, as it has not been tested.

6.3 Experiment 2: Prosody of semantic and pragmatic relations

In the research reported in Chapter 5, Mann and Thompson's (1988) list of 'subject matter' and 'presentational' relations was used to categorize rhetorical relations as either semantic or pragmatic (Sanders, 1992). They can also be described in terms of the 'source' of a coherence, i.e., a segment is related to another segment either semantically or pragmatically (Sanders, Spooren & Noordman, 1992). Whether a rhetorical relation is semantic or pragmatic depends on the kind of information readers use to make a coherent representation of two related segments. A relation is semantic if readers make coherent representations based on the content of propositions, for example, in '(s1) John is sleepy (s2) because he went to bed very late last night'. A relation is pragmatic if readers make coherent representations based on the illocutionary meaning of one of the segments or both segments, for example, in '(s1) John is sleepy. (s2) He looks very tired.' In the study reported in Chapter 5, no prosodic differences were observed between pragmatically and semantically related segments. However, there were several confounding factors as was explained in section 6.1. In the experiment described in this section, the effects of semantic and pragmatic relations on prosody were investigated in a controlled way. Text pairs are constructed containing an identical target sentence which is either semantically or pragmatically related to its preceding sentence.

6.3.1 Pretest: Construction and selection of text material

There were twenty target sentences. Each target sentence was included in two texts. In one of the texts, the target sentence was semantically related to its preceding sentence; in the other text, it was pragmatically related. Twenty pairs of texts resulted. A pretest of the text material was conducted to determine whether text manipulation succeeded in creating semantic and pragmatic interpretations of the target sentences. The pretest consisted of both semantic/pragmatic judgments and plausibility judgments. Semantic/pragmatic judgments were concerned with the extent to which target sentences were semantically or pragmatically related to the preceding sentences; plausibility judgments were concerned with the extent to which the target sentences followed their preceding sentences plausibly.

6.3.1.1 Text material

Twenty pairs of texts were constructed. Each pair was constructed in two conditions: in one version of the text, a target sentence was semantically related to its preceding sentence; in the other version, an identical target sentence was pragmatically related to its preceding sentence. Sign relations and generalizations were distinguished in both conditions. This distinction is explained below. The target sentence was preceded by two or three other sentences, and followed by one sentence. Each text was preceded by a context, i.e., a description of the situation in which speakers had to imagine they were involved before they read the text aloud. An example of a semantically related sentence is presented in (3); an example of a pragmatically related one is presented in (4). The target sentences are printed in bold; the preceding sentences to which the target sentences are related are printed in italics.

- (3)
semantic relation
(sign)
- Context:
A friend inquires about your housemate Alex. The friend knows that Alex has to take his driving-test today and wonders how it will go. You spoke to Alex this morning. He told you that he was nervous and that he was afraid of making mistakes because of that. You tell your friend what Alex told you.
- Text:
Alex has to check in at the driving school at two o'clock. *He is afraid of making stupid mistakes. **He is nervous.*** He will surely call this afternoon to tell us how it went."
- [Context:
Een vriend informeert naar je huisgenoot Alex. De vriend weet dat Alex vandaag zijn rij-examen moet afleggen en is benieuwd hoe dat zal gaan aflopen. Vanmorgen heb je Alex nog gesproken. Hij zei toen dat hij nerveus was en dat hij bang was dat hij daardoor vergissingen zou maken. Je vertelt je vriend wat je van Alex hebt gehoord.
- Tekst:
"Alex moet zich om twee uur bij de rijsschool melden. *Hij is bang dat hij domme fouten zal maken. **Hij is nerveus.*** Hij zal vanmiddag wel bellen om te vertellen hoe het is afgelopen."]
- (4)
pragmatic relation
(sign)
- Context:
You talk with a classmate about your common friend Piet. You both know that Piet has to give a presentation today. You saw him sitting in class, but you were not able to ask whether he dreaded for the presentation.
- Text:
I was not sitting near Piet. I did not talk to him. *I saw he was biting his nails all the time. **He is nervous.*** I hope that he is well prepared."]
- [Context:
Je praat met een klasgenoot over jullie gemeenschappelijke vriend Piet. Jullie weten dat Piet vandaag een presentatie moet houden. Jullie hebben hem wel in de klas zien zitten, maar konden hem niet vragen of hij tegen de presentatie op zag.
- Tekst:
"Ik zat een eindje bij Piet vandaan. Ik heb hem niet gesproken. *Ik zag dat hij de hele tijd op zijn nagels zat te bijten. **Hij is nerveus.*** Ik hoop voor hem dat hij zich goed heeft voorbereid."]

The contexts were designed so as to evoke a semantic or pragmatic interpretation of the target sentence. In the semantic condition, it was said in the context that the speaker was familiar with the fact mentioned in the target sentence; in the pragmatic condition it was clear that the speaker was not sure of the fact mentioned in the target sentence. Speakers, therefore, would present the target sentence in the semantic condition as a known fact, whereas they would present it in the pragmatic condition as their own conclusion. For example, based on the context in (3), the speaker is well informed about Alex's feelings, and, therefore, it is a semantic relation. The target sentence in (3) can be paraphrased as 'because he is nervous'. Based on the context in (4), the speaker has reasons to believe that Piet is nervous, but can not be totally sure about that. The target sentence in (4) can be paraphrased as 'so he must be nervous'.

For a good comparison of their prosodic characteristics, the target sentences had to be identical in the two conditions. Therefore, in the semantic condition, connectives were avoided (such as *therefore, because*); in the pragmatic condition, modal words which would point to a

pragmatic interpretation (such as *would, could, sure, I think, it may be the case that*), and connectives (such as *so, since*) were avoided. Because of the need to leave out linguistic markers, using a context seemed to be the only way to evoke a pragmatic or semantic interpretation of the target sentence.

The texts were written in the first person singular, presenting a speaker's perspective. The contexts and texts were constructed without gender-specific characteristics because they had to be read aloud by both male and female speakers, i.e., some descriptions of contexts were stated in terms of 'you are telling your partner' instead of 'your wife' or 'your husband'. Both semantic and pragmatic relations were causal relations. The direction of causality was held constant in all texts: in the semantic condition, the target sentence presented the cause and its preceding sentence the result; in the pragmatic condition, the target sentence presented the conclusion and its preceding sentence the fact on which the conclusion was based.

The target sentences in the pragmatic condition contained a conclusion about the non-perceptible state of mind of a person mentioned earlier. For the pragmatically related target sentences, two types of conclusions were distinguished, namely, sign relations and generalizations. In a sign relation, the conclusion was based on one observation; in a generalization, the conclusion was based on two observations. For the pragmatic condition, example (4) is a sign relation and example (5) is a generalization.

(5)
pragmatic relation
(generalization)

Context:

A builder has to be contracted for the construction of a new office premises. Your boss asks your opinion because you have much contact with builders. Your boss has heard something about a particular builder, and asks you whether you think he is suitable. You cannot be sure whether the builder is good. You base your judgment on things you have heard about him here and there.

Text:

I don't know whether he is the most suitable man for that order. *I heard that he has no office and no fixed personnel, and also that he went bankrupt once. He is unreliable.* I would not give him that order.

[Context:

Voor de bouw van een nieuw kantoorpand moet een aannemer worden gecontracteerd. Omdat jij veel contacten onderhoudt met aannemers vraagt je baas naar je mening. Hij heeft gehoord over een aannemer en vraagt aan jou of jij hem geschikt vindt. Je kunt niet met zekerheid zeggen of de aannemer een goede partij is. Je baseert je oordeel op dingen die je hier en daar eens over de man hebt gehoord.

Tekst:

Ik weet niet of hij de meest geschikte man is voor die opdracht. *Ik heb gehoord dat hij geen kantoor en geen vast personeel heeft en ook dat hij al een keer failliet is gegaan. Hij is onbetrouwbaar.* Ik zou hem geen opdracht geven.]

In the semantic condition, these two types of relations were adopted, too. The 'sign relation' in the semantic condition consisted of one observation and an explanation, whereas the 'generalization' consisted of two observations and an explanation. For the semantic condition, example (3) is a sign relation and example (6) is a generalization. Ten pairs of texts contained sign relations and ten pairs of texts contained generalizations.

(6)
semantic relation
(generalization)

Context:

You have heard from a business acquaintance that nobody wants to do business with Herman anymore. It appears that Herman does not keep his agreements and, therefore, he can not be relied on. A colleague knows nothing about these things, and asks you how Herman is doing. You tell your colleague what you have heard from your business acquaintance.

Text:

Herman's company is not going well. He has not kept his agreements. *Nobody wants to do business with him anymore.* **He is unreliable.** He has difficulties too in keeping his personnel.

[Context:

Van een zakenkennis heb je gehoord dat niemand meer zaken wil doen met Herman. Herman blijkt nooit zijn afspraken na te komen en is daardoor erg onbetrouwbaar. Een collega weet hier niks van en vraagt hoe het zit met Herman. Je vertelt je collega wat je van je zakenkennis hebt gehoord.

Tekst:

Het bedrijf van Herman loopt niet goed. Hij heeft zich nooit aan afspraken gehouden. *Er is niemand die nog zaken met hem wil doen.* **Hij is onbetrouwbaar.** Hij heeft ook moeite om zijn personeel vast te houden.]

6.3.1.2 Judges

Sixteen experts in the field of discourse studies participated in the semantic/pragmatic test. Most were affiliated with the Discourse Studies Group at Tilburg University; the others were affiliated with the Faculties of Arts at Utrecht University, Nijmegen University, Vrije Universiteit Amsterdam, and the University of Louvain-la-Neuve. They were all familiar with the theory of rhetorical relations in discourse and the semantic-pragmatic distinction. Forty-four students of the Faculty of Arts at Tilburg University participated in the plausibility test. They were not paid for the task.

6.3.1.3 Procedure

The semantic/pragmatic test consisted of twenty pairs of texts in the semantic and pragmatic conditions, the semanticity or pragmaticity of which had to be judged. The text pairs were distributed over two lists such that the semantic condition of a text pair did not co-occur with the pragmatic condition of the pair. Each list consisted of five sign relations and five generalizations in both conditions. Each judge received one of those lists, so that each text was judged by eight persons. The semantic-pragmatic distinction and the function of the context were briefly explained in the introduction. Judges had to indicate the type of rhetorical relation between the two sentences on a five-point scale ranging from 'strong semantic' to 'strong pragmatic'. The instructions are presented in Appendix F. The target sentences were printed in bold and their preceding sentences in italics to indicate that the judges had to examine the rhetorical relation between those two sentences. Judges were not allowed to communicate with each other about the task. The task was self-paced.

Twelve pairs of texts remained on the basis of the results of the semantic/pragmatic test, the results of which are explained below. The plausibility test contained these twelve pairs. They

were split up into two lists: half of the semantic conditions were combined with the other half of the pragmatic conditions, so that the semantic condition of a text pair never co-occurred with the pragmatic condition of that pair. Each judge received one of those lists, so that each text was judged by twenty-two persons. They had to indicate to what extent target sentences followed their preceding sentences plausibly on a five-point scale ranging from ‘very implausible’ to ‘very plausible’. The instruction were the same as those for the experiment on causal and non-causal relations presented in Appendix E. Target sentences and preceding sentences were printed in bold to indicate that the plausibility between those two sentences had to be judged. Judges were not allowed to communicate with each other about the task. The task was self-paced.

6.3.1.4 Results

The selection of texts to be used in the main study was based on the semantic/pragmatic judgments. These judgments were considered more important than the plausibility judgments, because the semantic or pragmatic relation between a target sentence and its preceding sentence was the independent variable in the main study. The distribution of the semantic/pragmatic judgments and the selection procedure of the texts to be used in the main study are first described. The results of the plausibility test are then described.

Table 6.5 presents the judgments of the semantic and pragmatic relations split into sign relations and generalizations. Scores are the number of judgments of the extent to which the relations in the text were either semantic or pragmatic relations as a function of the intended semantic and pragmatic relations.

Table 6.5 Distribution of semantic and pragmatic judgments in relation to the intended semantic and pragmatic relations and their subtypes (each column: n = 80)

judgments	semantic relations as intended		pragmatic relations as intended	
	sign relations	generalizations	sign relations	generalizations
‘strong pragmatic’	0	5	61	45
‘weak pragmatic’	3	5	15	25
‘unclear’	4	5	3	4
‘weak semantic’	10	15	1	5
‘strong semantic’	63	50	0	1

Semantically related sentences were predominantly scored as semantic and pragmatically related sentences as pragmatic.

The judgments were then transformed into scores ranging from 1 to 5, such that high scores reflect high correspondences to the intended rhetorical relations; for example, score 5 was assigned to ‘strong semantic’ judgments if the rhetorical relation was intended to be semantic, and to ‘strong pragmatic’ judgments if the rhetorical relation was intended to be pragmatic; score 4 was assigned to ‘weak semantic’ judgments if the rhetorical relation was intended to be semantic, and to ‘weak pragmatic’ judgments if the rhetorical relation was intended to be

pragmatic, and so forth. The mean score was 4.44 (sd: 0.40) for the twenty semantic relations and 4.52 (sd: 0.35) for the twenty pragmatic relations. Judgments of semantic relations did not differ from judgments of pragmatic relations ($t(38)=0.71$, $p=.48$): they did not differ with regard to the extent they were intended. Over both semantic and pragmatic relations, the mean score for the twenty sign relations was 4.64 (sd: 0.34) and, for the twenty generalizations, 4.33 (sd: 0.35). Judgments of sign relations were higher than those of generalizations ($t(38)=2.76$, $p<.01$).

The texts to be used in the main study were selected in two steps. First, when two or more judges scored the rhetorical relation between a target sentence and its preceding sentence as strong or weak pragmatic when it was intended to be semantic, and vice versa (score < 4), the text was removed. The other condition of the text pair was removed, too. Four pairs of texts were removed on the basis of this criterion. Second, when two or more judges indicated having serious doubts about the interpretation of the rhetorical relation, the text was removed, as was the other condition of the text pair. Another four pairs of texts were removed on the basis of this criterion. Twelve pairs of texts remained, eight of which were sign relations and four of which were generalizations. After this selection procedure, the difference between sign relations and generalizations was tested again. Judgments of sign relations no longer differed from those of generalizations (means: 4.66 and 4.46, respectively; $t(22)=1.20$, $p=.24$). The mean score over all selected texts was 4.59. This means that the judges interpreted the rhetorical relations almost unanimously as either semantic or pragmatic. The twelve pairs of texts are presented in Appendix H.

A plausibility test was performed on this final set of twelve pairs of texts. The plausibility of the semantically related target sentences was 3.80 (sd: .39), with a range from 3.13 to 4.33; the plausibility of the pragmatically related target sentences was 3.54 (sd: .50), with a range from 2.79 to 4.23. The plausibility was not different for the two conditions ($t(22)=1.42$, $p=.17$). The selected texts were considered clear manipulations of semantic and pragmatic relations.

6.3.2 Main study: Prosodic realization of semantic and pragmatic relations

6.3.2.1 Speakers

Twenty-four speakers participated in the experiment. They were native speakers of Dutch. Because a broad experience of reading was required, students were chosen. They were from the faculties of arts and psychology of Tilburg University. There were 7 male and 17 female speakers. Eight had participated also as speakers in the experiment on causal and non-causal relations. The mean age was 21.8 years. The speakers were not informed about the goal of the experiment. They were paid for their participation.

6.3.2.2 Procedure

The procedure of reading the texts aloud was the same as for the experiment on causal and non-causal relations. The function of the context was clearly explained: the speakers were encouraged to empathize with their roles as if they were actors in a play, and to prepare themselves thoroughly to read the texts. Recordings of the texts were preceded by two training texts. The

experiment was run in two blocks of twelve texts. Recordings were started again from the beginning of a text when the speech was not fluent. This happened on average twice for each speaker. The reading sessions lasted about twenty minutes.

6.3.2.3 Speech material

The prosodic characteristics were the same as those in the study reported in Chapter 5: duration of the preceding pause, the F0-maximum, and articulation rate of the target sentence. In addition, pause duration following the target sentence was measured, because it was expected that pauses preceding and following a target sentence might be related in some way. In addition to the F0-maximum, the mean pitch range of the whole target sentence was also measured, because a change in the pitch contour of the target sentence was expected between both conditions, perhaps resulting in a different mean pitch range, rather than a different F0-maximum.

6.3.2.4 Results

Table 6.6 presents the mean pause durations, F0-maximum, mean pitch range, and articulation rate for semantic and pragmatic relations.

Table 6.6 Prosodic characteristics of semantic and pragmatic relations

	semantic	pragmatic	F ₁	F ₂
pause preceding (in milliseconds)	352	449	***	***
pause following (in milliseconds)	458	434		
F0-maximum (in hertz)	224.1	233.6	***	**
mean pitch range (in hertz)	174.5	179.1	***	*
articulation rate (in phonemes per second)	16.7	16.6		

Note: * p<.05; ** p<.01; *** p<.001

F₁ and F₂ analyses of variance with repeated measurements were run. The independent variable was type of rhetorical relation (two levels: pragmatic, semantic). Speaker was the random variable in the F₁ analysis; target sentence in the F₂ analysis. The order of the texts was included as a between-groups factor (two levels: text 1-24, text 24-1). In none of the analyses was order a factor of interest; therefore, no results are reported.

The pauses preceding target sentences were on average 97 milliseconds longer in the pragmatic condition than in the semantic condition (F₁ (1, 23) = 29.96, p<.001, η^2 = .57; F₂ (1, 11) = 25.40, p<.001, η^2 = .70).

The pauses following target sentences did not differ for pragmatic and semantic relations (F₁ (1, 23) = 3.09, p=.09, η^2 = .12; F₂<1).

The F0-maxima of pragmatically related target sentences were on average about 10 hertz higher than those of semantically related target sentences (F₁ (1, 23) = 17.98, p<.001, η^2 = .44; F₂ (1, 11) = 7.76, p<.05, η^2 = .41). The mean pitch of pragmatically related target sentences was

on average about 5 hertz higher than that of semantically related target sentences ($F_1(1, 23) = 13.13, p < .001, \eta^2 = .36$; $F_2(1, 11) = 6.24, p < .05, \eta^2 = .36$).

Articulation rate was not affected by type of rhetorical relation ($F_1 < 1$; $F_2 < 1$).

For pauses at both positions, the same analysis of variance was performed as in the experiment on causal and non-causal relations. There were two within-group factors: Location of pause (two levels: preceding, following) and Rhetorical relation (two levels: semantic, pragmatic). There was an interaction between type of rhetorical relation and location of pauses ($F_1(1, 23) = 26.36, p < .001, \eta^2 = .53$; $F_2(1, 11) = 8.91, p < .05, \eta^2 = .45$). In pragmatic relations, pauses preceding and following the target sentences were equal (both F_1 and F_2 : $F < 1$); in semantic relations, pauses preceding target sentences were shorter than those following target sentences ($F_1(1, 23) = 19.60, p < .001, \eta^2 = .47$; $F_2(1, 11) = 15.53, p < .01, \eta^2 = .59$).

Table 6.7 presents for each speaker the difference scores of the prosodic characteristics between semantic and pragmatic relations. The scores of semantic relations were subtracted from the corresponding scores of pragmatic relations. Articulation rate was not included in the table, because it was not affected by type of rhetorical relation. In addition to the analyses of variance, differences between semantic and pragmatic relations were tested using Wilcoxon-matched-pairs tests.

Table 6.7 For each speaker difference between semantic and pragmatic relations in pause duration (in milliseconds), F0-maximum, and mean pitch (in hertz)

	preceding pause	F0-maximum	mean pitch
speaker 1	+ 5	+ 2.5	+ 1.0
speaker 2	+ 45	+ 0.4	- 1.0
speaker 3	+ 139	+ 20.3	+ 7.7
speaker 4	+ 6	+ 11.4	+ 6.7
speaker 5	+ 154	+ 1.7	+ 1.0
speaker 6	+ 68	+ 10.3	+ 6.2
speaker 7	+ 188	- 0.1	- 8.3
speaker 8	+ 105	+ 15.4	+ 8.2
speaker 9	+ 47	+ 12	+ 2.4
speaker 10	+ 75	+ 34.2	+ 2.4
speaker 11	+ 132	+ 5.5	+ 5.1
speaker 12	- 39	- 8.1	- 2.3
speaker 13	+ 74	+ 11.5	+ 1.5
speaker 14	+ 116	+ 3.2	+ 2.5
speaker 15	+ 123	+ 6.8	+ 8.7
speaker 16	+ 217	+ 0.6	- 1.6
speaker 17	- 58	+ 3.6	+ 2.3
speaker 18	+ 312	+ 33.8	+ 14.7
speaker 19	+ 31	+ 1.3	+ 1.2
speaker 20	+ 206	+ 8.3	+ 11.2
speaker 21	+ 69	- 4.1	- 0.7
speaker 22	+ 19	+ 14.8	+ 16.7
speaker 23	+ 107	+ 28.6	+ 17.8
speaker 24	+ 187	+ 16.2	+ 8.3
pragmatic > semantic	N = 22	N = 21	N = 19
pragmatic < semantic	N = 2	N = 3	N = 5

Twenty two speakers produced longer pauses preceding pragmatically related sentences than preceding semantically related ones ($z=3.91$, $p<.001$). Twenty-one speakers produced higher F0-maxima in target sentences in pragmatically related sentences ($z= 3.66$, $p<.001$). Nineteen speakers produced a higher mean pitch in pragmatically related sentences ($z=3.22$ $p<.01$). Table 6.8 presents for each text the difference scores of the prosodic measurements between semantic and pragmatic relations.

Table 6.8 For each text differences between pragmatic and semantic relations in preceding pause duration (in milliseconds), F0-maximum, and mean pitch range of the target sentence (in hertz)

	preceding pause	F0-maximum	mean pitch range
text 1	+ 76	+ 2.7	+ 1.6
text 2	+ 179	- 2.5	+ 2.7
text 3	+ 53	- 11.1	- 5.9
text 4	+ 165	+ 8.9	- 4.9
text 5	+ 85	+ 8.6	+ 3.6
text 6	+ 63	+ 3.1	+ 2.9
text 7	+ 130	+ 7.1	+ 2.3
text 8	+ 132	+ 30.8	+ 15.4
text 9	- 31	+ 8.6	+ 13.2
text 10	+ 103	+ 10.5	+ 8.0
text 11	+ 18	+ 21.6	+ 7.5
text 12	+ 192	+ 26.8	+ 8.5
pragmatic > semantic	N = 11	N = 10	N = 10
pragmatic < semantic	N = 1	N = 2	N = 2

In eleven texts, pauses preceding target sentences were longer in the pragmatic condition ($z=2.90$, $p<.01$). The effect was consistent for both the F0-maximum and mean pitch range: in ten of the twelve texts, F0-maximum and mean pitch range were higher in the pragmatic condition (F0-maximum: $z=2.28$, $p<.05$; mean pitch range: $z=2.04$ $p<.05$).

6.3.2.5 Discussion and conclusion

The pauses preceding pragmatically related target sentences were longer than those preceding semantically related target sentences. The longer pause durations reflect that two pragmatically related sentences cohere less than two semantically related sentences. This may be explained by the change of the speaker's perspective in the pragmatic condition because the description of events has been interrupted by a concluding remark. In the corpus study described in Chapter 5, the pauses preceding the pragmatically related segments were not longer than those preceding the semantically related segments. This experiment, however, may be considered more sensitive to the effect of rhetorical relations on prosody than the corpus study. Because the scores of the pretest were high, the experimental material may be regarded as an adequate operationalization of both kinds of relations. In addition, both kinds of rhetorical relations were operationalized more precisely and uniformly in the experiment than in the corpus study. Therefore, the results of the experiment have to be taken as more conclusive than the results of the corpus study.

There was no effect of type of rhetorical relation on pause duration following the target sentences. However, the patterns of pauses preceding and following target sentences differed for the two relations. In pragmatic relations, the pauses preceding and following the target sentences were equal in length, whereas in semantic relations, the preceding pauses were much shorter than the following pauses. Because the pause between the target sentence and its preceding sentence

was very short in the semantic condition, whereas the pause following the target sentence was much longer, it seems that the closeness of the semantically related sentences reflects the closeness of the described events in the world. In the pragmatic condition, the pauses between the target sentences and their preceding sentences were not shorter than the pauses following the target sentences, which may indicate that the observation(s) and the conclusion based on it were read aloud as though there was not a close relation between them. The pause patterns in the two conditions suggest that the relation between two semantically related sentences is stronger than the relation between two pragmatically related sentences. The longer pauses between pragmatically related sentences may be interpreted as a shift of perspective.

Pragmatically related sentences were realized with a higher F0-maximum and mean pitch range than semantically related sentences. These results for pitch range may reflect the same shift of perspective as mentioned above. A pragmatic relation indicates that writers interrupt their descriptions of actual events, for example, by drawing personal conclusions or making remarks. The raising of the pitch in the production of pragmatically related sentences seems to show that speakers recognize these sentences as being reflections of the writer's changed perspective, and present them as such when they read them aloud.

No effects on articulation rate were found in the experiment.

In the experimental text material, no linguistic means were present to mark semantic and pragmatic relations, such as connectives or other lexical indicators. The results show that, in spite of the lack of lexical signals, speakers mentally registered that the sentences cohere in different ways. Based on the high plausibility of the target sentences, it was assumed that neither semantic nor pragmatic relations suffered much from the lack of lexical marking. The use of a context would have compensated sufficiently for that. This assumption is supported by the fact that the plausibility scores of the target sentences did not correlate with prosodic cues in either condition. Table 6.9 presents the correlations between plausibility, on the one hand, and the preceding and following pauses, F0-maximum, mean pitch range, and articulation rate, on the other hand, of both semantically and pragmatically related target sentences.

Table 6.9 For semantically and pragmatically related target sentences correlations between plausibility scores and prosodic characteristics

		preceding pause	following pause	F0-maximum	mean pitch range	rate
semantic	(n=12)	-.29	-.33	.42	.09	.12
pragmatic	(n=12)	.01	.42	.37	.42	.11

The experiment must be replicated using lexical markers of the rhetorical relation in order to investigate whether the effects of the rhetorical relation on prosody remain the same in both conditions. The results of the experiment seem to show that, when no other lexical means are available to express the rhetorical relation, speakers use pause duration and pitch range to do so. Such an experimental approach could only be performed on the durations of the pauses preceding the sentences, and not on the F0-maxima of the sentences, because adding a lexical marker would affect the intonation pattern of the sentence, but not the duration of the preceding pause.

Finally, an informal and impressionistic observation in the speech material of the experiment on semantic and pragmatic relations is that the difference in the prosodic realization between semantic and pragmatic relations may partly be due to a difference in the whole intonation contour: the main accent in the segment seemed to shift. If so, such a shift would have a direct consequence for the location of the F0-maximum. The pattern of the whole intonation contour was not attended to in the present study, but it deserves closer attention in further research on the prosodic marking of rhetorical relations.

6.4 Conclusion and discussion

In the experiment on causal and non-causal relations, a consistent effect on mean pitch range was found. The effect of pause duration was significant only for speakers, and not for texts, which indicates that the text material in which causal and non-causal relations occur is important. In the experiment on semantic and pragmatic relations, consistent effects on pause duration, the F0-maximum, and mean pitch range were found. The distinction between semantically and pragmatically related sentences seems to be marked prosodically more clearly than the distinction between causally and non-causally related sentences. Earlier work on rhetorical relations (for example, Murray, 1997) showed that causally and non-causally related sentences do not differ in terms of continuity, whereas semantically and pragmatically related sentences do. Semantically related sentences are considered to be more continuous, and pragmatically related sentences more discontinuous, whereas this distinction cannot be made for causal and non-causal relations. The longer preceding pauses, and the higher F0-maximum and mean pitch range in pragmatically related sentences are expressions of the discontinuity between pragmatically related sentences. Pragmatic relations may, therefore, require stronger markers than other rhetorical relations. In the experiment, the speakers had to accentuate the discontinuity by prosodic means, because there were no other ways to do so. Strikingly, the construction of the pragmatically related sentences without using lexical signals, like modal verbs or explicitly adding the writer's perspective (*'I think...'*), was much more difficult than the construction of the other rhetorical relations between sentences. Lexical markers were felt to be more necessary in pragmatic relations than in other relations. In this experiment, the high plausibility scores indicate that the use of a context compensated sufficiently for the difficulty of recognizing the pragmatic relations.

The prosodic characteristics of pragmatic relations may be explained from both a speaker's and a listener's point of view. On the one hand, the speakers may have recognized the writer's shift of perspective in the pragmatic relations, and may have expressed that prosodically. On the other hand, the speakers may have accommodated the supposed need of listeners to get a clear understanding of what was read aloud to them. Whether listeners perceive the prosodic characteristics of rhetorical relations is not yet known. Therefore, perception experiments are needed using the same text materials as were used in this experiment.

7

Discussion

7.1 Conclusions

The studies described in this dissertation focus on the clarification of the relation between text structure and prosody. They provide empirical evidence for three research domains: first, the reliability and relevance of procedures for assigning text structure; second, the reliability and relevance of measurements of prosodic characteristics; third, the actual relation between texts and their prosody. The conclusions for these three domains of research are explicated in the following sections.

7.1.1 Analyses of text structure

The study described in Chapter 2 looked at the reliability of text structure analyses. The question to be answered was whether analysts come up with the same structure analyses of a text when they apply a particular procedure independently of each other. In the study, four natural texts were analyzed in four ways. Two intuition-based procedures were used, i.e., naive subjects indicated paragraph boundaries in the texts. In one procedure, the number of boundary markers was unlimited; in the other, it was restricted. Two theory-based procedures were also used: expert users analyzed the texts using Intention Based Analysis and Rhetorical Structure Theory. The reliability of the text structures which were resulting from these procedures was statistically evaluated. One of the intuition-based procedures and one of the theory-based procedures were found to be applied reliably. The restricted variant of the intuitive procedure was applied reliably. The inter-subject reliability was lower when subjects were free to decide how many boundaries to indicate in a text. The theory-based procedure, IBA, was conspicuously less reliably applied than was RST. Using the explicit relation definitions in RST, resulting in an in-depth processing of a text by an analyst, provided more reliable text structures. We showed that, first, RST could be applied reliably by a single trained person to obtain the hierarchical structures needed in this dissertation; second, it provided the multilayered hierarchical structures of the texts that were needed; third, it gave specific information on the rhetorical relations existing between parts of the texts. These characteristics of RST provided a solid basis for investigating the relation between text structure and its prosodic correlates.

In the study described in Chapter 4, the same four texts were used as in the study described in Chapter 2. The texts were originally broadcast on the radio, and, therefore, the prosodic realizations were available. The study described in Chapter 4 was explorative in more than one way. The average scores for hierarchical level of the six RST analyses were used instead of the scores for hierarchical level of one particular RST analysis. This led to methodological problems. For example, because the average scores for the boundaries were used, some numeric levels no longer occurred. The study was also explorative in that various ways of quantifying the hierarchical structures were tried. A top-down procedure, a bottom-up procedure, and a compromise between these two, a symmetrical procedure, were used. The procedures did not provide fundamentally different results: they varied only with regard to the sample of segments which were involved in the analyses. The symmetrical procedure was adopted for use in Chapter 5.

7.1.2 Measurements of prosody

In most studies of this dissertation, a fixed set of prosodic parameters were used to investigate their marking function for hierarchy in text structures. The parameters were pause duration, pitch range, and articulation rate. Other prosodic parameters may also have been useful for marking text structure, but in earlier research the three parameters selected had been found to be related to some aspects of text structure.

Pause duration and articulation rate could be observed without difficulties. For pitch range, however, a relevant measurement was more difficult to find since the pitch contour of a whole segment had to be characterized by one particular score for pitch range. The use of declination lines as a characterization of the pitch contour of a whole segment seemed to be more appropriate than the use of one particular value of the contour, the F0-maximum. Declination lines can be computed automatically using linear regression lines, but a disadvantage is that they do not reflect the perceptually relevant peaks and valleys of a pitch contour. They can only be taken into account by human judges. Linguistic knowledge is also required for the measurement of the F0-maximum. The study described in Chapter 3 was performed to investigate the reliability of the measurements of the declination lines and the measurements of the F0-maximum. Five phoneticians determined the declination lines and the F0-maxima in forty pieces of speech. The reliability of the measurements of the F0-maxima was considerably higher than that of the measurements of the declination lines. The judges agreed strongly about the value of the F0-maximum, especially in relative terms. It can be measured automatically in an adequate way, and the score was found to be independent of the length of an utterance. For the speech used in this dissertation, i.e., prepared read-aloud speech, we showed that the F0-maximum is an adequate measure for characterizing the pitch range.

7.1.3 The relation between text structure and prosody

In the study described in Chapter 4, two approaches to relating the scores for hierarchy and the prosodic measurements, i.e., a relative approach and an absolute approach, were explored. In the relative approach, the prosodic features of segments were compared pairwise: either the prosodic features of adjacent segments in the text were compared pairwise, or the prosodic features of dominating segments and dominated segments were compared pairwise. In this approach, the level scores were considered ordinal ones, i.e., the boundaries of a pair were characterized as 'higher' and 'lower', and then related to the means of the prosodic realizations of the higher and lower boundaries. In the absolute approach, the levels of the hierarchical structures, defined on an interval scale, were directly related to the prosodic features of the segments at these levels. The results of the relative and absolute approaches were very similar: the duration of the pause preceding a segment and the F0-maximum of a segment were found to be affected by the hierarchical level in the text structure. The higher the level of a segment in the hierarchy, the longer the preceding pause and the higher the F0-maximum. No such trend was found for articulation rate. Initially, the relative and absolute approaches were distinguished because the perspective on hierarchical structure in texts in relation to prosody was different in both approaches. Although the results with regard to prosodic realization did not differ, distinguishing

the approaches was still regarded relevant for the relation between hierarchy and prosody. Therefore, they were maintained in the research described in Chapter 5.

In the study reported in Chapter 5, a corpus of twenty texts was used, consisting of one text type: descriptive, narrative texts. The texts were first analyzed using Rhetorical Structure Theory: hierarchical levels were determined using the symmetrical procedure of scoring; nuclei and satellites were determined; and rhetorical relations between text parts were determined. They were then read aloud by twenty different speakers. Finally, the prosodic features of the segments were measured, and related to the relative and absolute hierarchical levels, the nuclearity, and the rhetorical relations in the texts. The hierarchical levels of the texts were marked by pause duration and the F0-maximum. The pattern found was the same as that in the study described in Chapter 4: pauses at higher boundaries had longer durations than pauses at lower boundaries, and the F0-maxima of the segments following the higher boundaries was higher than the F0-maxima of the segments following the lower boundaries. Nuclearity was marked by articulation rate: nuclei were read at a slower rate than satellites. Causality was marked by pause duration and articulation rate: preceding pauses were shorter for causally related segments than for non-causally related ones, and the articulation rate was higher for causally related segments than for non-causally related ones. Semanticity did not affect prosodic features.

The prosodic realization of rhetorical relations was investigated more closely in the two experiments described in Chapter 6. One experiment dealt with causal and non-causal relations; the other dealt with semantic and pragmatic relations. In the experiments, identical target sentences were constructed which were either causally and non-causally or semantically and pragmatically related to a preceding sentence. The target sentence and its preceding sentence were part of a short text. More than twenty speakers read these texts aloud. The speakers realized causally related target sentences with somewhat higher mean pitch range than they did non-causally related target sentences. The speakers realized pragmatically related target sentences with longer preceding pauses and higher pitch range than they did semantically related target sentences.

The two experimental studies on the prosodic realization of specific relation types pose a number of questions. The corpus study reported in Chapter 5 and the experimental study reported in Chapter 6 gave some opposite results. In the natural text material, causally related segments had shorter preceding pauses and faster articulation rates than non-causally related segments, whereas, in the manipulated text material, causally related segments had a higher mean pitch range than non-causally related segments, and no significant differences were found for the preceding pause and articulation rate. It was assumed that the lexical markedness of the rhetorical relations could explain these differences. In the experiments described in Chapter 6, lexical markers were intentionally removed to keep the target sentences in the two conditions identical. In these constructed texts, the absence of lexical markers may have led to an increase of the importance of the contribution of prosody.

The implications of the results of the studies reported in Chapters 4, 5, and 6 follow from the evaluation of the relation between text structure and prosody. They are closely connected with the objectives of this dissertation, i.e., contributing to the improvement of automatically generated texts, and the theoretical modeling of human text production. They are explained in the next two sections.

7.2 Implications for text-to-speech systems

Speakers prosodically realized the hierarchical levels of text structure most notably in the durations of the pauses preceding segments, and in the F0-maximum of the segments. The structural marking of hierarchical structure is missing in current text-to-speech systems. These systems only mark boundaries between paragraphs with pauses that are somewhat longer than the pauses at boundaries within the paragraph. In some cases, they also raise the F0-maximum of sentences following a paragraph boundary. The results of the studies reported in Chapters 5 and 6 show that more fine-tuned adjustments are needed in text-to-speech systems in several respects on the basis of both the level at which a specific segment figures in the text structure and the specific characteristics of its rhetorical relation.

To demonstrate how such adjustments can be implemented, estimates are computed for the prosodic parameters that showed a relation with text structure most conclusively: the pause preceding a segment and the F0-maximum of a segment. These estimates are derived from linear regression analyses of the data obtained for the speakers participating in the studies reported in Chapters 5 and 6.

For the prosodic marking of text structure, level scores are computed using the linear trend determined for the speaker with the best predictive power, that is, the one who contributed the highest correlation in Table 5.8 (see Chapter 5). For the pause preceding a segment, speaker 11 best predicted the hierarchical level ($R^2 = .55$)¹; the standard scores are computed using the regression formula $z = .59 * Level - 1.32$. For the F0-maximum of a segment, speaker 1 best predicted the hierarchical level ($R^2 = .65$); the standard scores are computed using the regression formula $z = .54 * Level - 1.36$. For five levels, standard scores are computed for pause duration and the F0-maximum. They are presented in Table 7.1. For instance, a segment at level 2 in the text structure has a standard score of -0.13 for preceding pause and a standard score of -0.29 for the F0-maximum.

To convert these standard scores into raw scores for actual use in text-to-speech systems, prosodic estimates are computed, for a male and a female speaker, using the data of all speakers participating in the corpus study reported in Chapter 5. The results were presented in Table 5.3. The pause preceding a segment lasted, on average, 917 milliseconds ($sd = 426$) for males and 801 milliseconds ($sd = 374$) for females. The F0-maximum reached, on average, 169 hertz ($sd = 29$) for males and 281 hertz ($sd = 34$) for females. In general, women took shorter pauses between segments. Each of the standard scores is transformed into a raw score by adding to the relevant mean the product of the standard score and its standard deviation. For instance, it can be read from Table 7.1, that a segment at level 2 in the text structure corresponds with a pause of 862 milliseconds for males, and with a F0-maximum of 271 hertz for females.

The RST analyses reported in Chapter 5 had levels ranging from 1 to 10. To evaluate the relation between text structure and prosody, these ten levels were reduced to five classes. Long texts contain more than five levels and, ideally, text-to-speech systems would need estimations for more than five levels of text structure. It is not realistic, though, to expect people working on text-to-speech systems, being mostly engineers, to be able to distinguish this subtlety.

¹ R^2 is the squared Pearson correlation between the two variables involved.

Furthermore, one may question the relevance of such a sophistication: a text-to-speech system with a prosodic marking of at least five textual levels may be expected to improve its output optimally.

Estimates for the prosodic marking of type of relation were made for the results of the experiments reported in Chapter 6. These were the data in Table 6.1 for causal and non-causal relations, and the data in Table 6.6 for semantic and pragmatic relations. Causally related segments differ from non-causally related segments in their mean pitch range by 3 hertz. Semantically related segments differ from pragmatically related segments in their preceding pause duration by 97 milliseconds, in their F0-maximum by 9 hertz, and in their mean pitch range by 5 hertz. The perceptual relevance of the differences in the pitch-range parameters is negligible for the relation types. Therefore, no estimates are made for the pitch range parameters. Pause duration is different for semantic and pragmatic relations. By either subtracting or adding half of the values, the raw score given for level in Table 7.1 can be adjusted for type of relation. For instance, for a pragmatic relation at level 4, the pause of a female speaker in a text-to-speech system was estimated at 1247 milliseconds ($1197 + 50$) and, for a semantic relation, at 1147 milliseconds ($1197 - 50$).

When the structure of a text is available (Di Christo, Auran, Bertrand, Chanet, & Portes, 2002; Carlson, Marcu, & Okurowski, 2003), hierarchical levels and semanticity can be adjusted in this way.

Table 7.1 Estimates for adjusting pause duration and the F0-maximum to hierarchical level, type of rhetorical relation, and gender in generated speech

			preceding pause duration			F0-maximum of segment		
			standard score	raw score (in msec)		standard score	raw score (in hertz)	
				male	female		male	female
Level in text structure	5	highest	1.65	1620	1418	1.34	208	327
	4		1.06	1369	1197	0.80	192	308
	3		0.46	1113	973	0.26	177	290
	2		- 0.13	862	801	- 0.29	161	271
	1	lowest	- 0.73	606	528	- 0.83	145	253
Type of relation	semantic:	subtract 50 msec						
	pragmatic:	add 50 msec						

7.3 Beyond the limitations of this research

The study of pauses has a long tradition in psycholinguistic research. In most cases, pausing in speech has been considered a signal of planning activity. In fact, the possibility that pauses may also have a marking function has been considered disturbing. Consider, for instance, the following remark by Harley (2001: 376): “There are a number of problems with pause analysis (...) It is possible that speakers deliberately (though perhaps usually unconsciously) put pauses

into their speech to make the listener's job easier". The aspect that speakers put pauses in their speech for other reasons than planning was the focus of attention in this dissertation. It was necessary to control explicitly for the influence of any planning activity, be it conceptual, phonological, syntactic, or lexical. It was a methodological decision to use prepared read-aloud speech in all studies. Before they read the texts aloud, speakers took a close look at their content and organization. This procedure gave the speakers the opportunity to prepare their speech delivery maximally, ensuring that they would adjust their prosody with regard to the text as it was, and would not be influenced by planning activities. This mode of speaking is usually called 'diction': the ways of making speech expressive as, for instance, when reading aloud a book or the news, or giving a lecture or a speech. In all these modes, speakers are mentally prepared regarding the content of their speech; they only have to concentrate on ways of expressing it to correspond maximally with their intentions. As a consequence, the results of the studies described in this dissertation can not be generalized to spontaneous speech. As far as there is a relation between the text-structural features investigated in this dissertation and their prosodic realizations in spontaneous speech, the text-structural features in spontaneous speech are far more difficult to describe.

The studies described in this dissertation concerned solely the acoustical analyses of the speech material, and not perceptual analyses. The differences in the pitch-range parameters for the relation types found in the study described in Chapter 6 were so small that they would certainly not be perceived by listeners. Speakers, however, were consistent in their realizations of pitch range. Unlike this consistency in production, speakers may not have intended to make the listener's job easier. To test whether speakers deliberately put pauses into their speech to make the listener's job easier, the perceptual relevance of the proposal to adjust a text's prosody has to be examined. As labeled hierarchical structures are available for the texts used in the study described in Chapter 5, one of these texts could be selected for prosodic manipulation in synthesized speech. The manipulation may be done using various experimental conditions. Listeners would have to indicate the extent to which they perceive that the text sounds natural. For example, in a 'constant condition', the segments of the presented text would not differ prosodically: the duration of the preceding pauses and the F0-maxima would have to be held constant for all segments. For male speakers, this would amount to an average preceding pause of 917 milliseconds and an average F0-maximum of 169 hertz for all segments, and for female speakers, to an average preceding pause of 801 milliseconds and an average F0-maximum of 281 hertz for all segments of the text. In the 'natural condition', the prosodic parameters of the segments of the text would be adjusted for hierarchical level and rhetorical relation as indicated in Table 7.1. For male and female synthesized speech, the adjustments would have to be made accordingly. A comparison between the 'constant' and the 'natural' version would make clear what contribution text prosody makes to the naturalness of synthesized speech. This comparison would not rule out a simpler explanation, i.e., it is not so much exactly where the specific parameter values are realized, but simply the fact that the prosodic realization of the text contains variation that is important. Therefore, a 'reversed condition' could be used, in which the prosodic parameters of the segments would be adjusted for hierarchical level and rhetorical relation in an order that is completely the reverse of that in Table 7.1. For instance, a pause at level 5 would

be assigned the value of a pause at level 1. If only variation matters, the 'natural' and 'reversed' versions should be evaluated equally.

Concerning the possible perceptual relevance of the results found in the production studies described in this dissertation, the length of the texts used in relation to the prosodic realizations of the hierarchical levels of the texts may also be discussed. In the studies described in this dissertation, texts containing about thirty segments were considered long texts providing hierarchical structures in which between five and ten levels could be distinguished. The relation between the levels and the prosodic realizations has been demonstrated conclusively. Chapters of books, or sections of chapters, however, are longer texts which provide even deeper hierarchical structures. Further reasoning of the results found may lead to pause durations of three minutes at boundaries between the chapters of a book. In the studies described in this dissertation, the limitations of the prosodic realizations of hierarchy were not indicated.

The results of the studies described in Chapter 5 and Chapter 6 seem to show that prosody and lexical marking might be related in some way. In an additional experiment, it may be demonstrated whether prosody functions independently of lexical marking. For example, speakers could read aloud causally related sentences in two conditions: a condition in which the sentences are related using a causal connective, and a condition in which they are related without using a causal connective. If prosody functions independently of lexical marking, then the prosodic features will be equal. If there is a form of trade-off between the prosodic and the lexical marking of text structure, prosodic differences will show up. In such an experiment, only the pause durations between the sentences would have to be measured, because they are not influenced by the markedness of the second sentences. It is different for the F0-maximum, because the addition of a lexical marker would change the features of the pitch contour. This would make it impossible to compare the lexically marked and lexically unmarked sentences.

The so-called 'diction' of language use raises the question of proficiency: are speakers equally competent in reading aloud? The results of the study in Chapter 5 showed large individual differences for the prosodic marking of structural aspects of text. Some speakers mirrored the structures of texts prosodically, whereas other speakers hardly showed any relation between text structure and prosody. Two explanations for these differences are possible. First, some speakers are less proficient orators than others and, therefore, produce more monotonous speech. They simply do not know how to realize a text structure prosodically. Second, some speakers are less proficient readers than others. They have more difficulties in recovering the structure of a text, and, therefore, they fail to notice the possibilities for prosodic marking. Further research can disentangle these two explanations. Since some people may be unaware of the structural properties of the texts, the speakers would be given additional tasks to prepare them more intensively for the reading-aloud session. Such tasks could be derived from the intuition-based text analyses. For instance, to make speakers more aware of text-structural features, they could be instructed to put markers at important boundaries, or to select those sentences that are central in the text (i.e., that are high in the hierarchy), or to cross out sentences of minor importance in the text (i.e., that are low in the hierarchy). Two questions could then be answered. First, do people realize text structure in a more subtle way following extended preparation than they would

following more passive preparation completely by themselves? Second, do individual differences between speakers in the prosodic marking of text-structural aspects decrease following extended preparation?

The theoretical modeling of human text production may be adjusted using the results in consequence of the studies described in this dissertation. Hierarchy in text structure is reflected by prosody, as well as nuclearity, causality, and semanticity. Whether readers of texts make a mental representation of hierarchy in a relative or in an absolute way is still to be investigated. The assumption underlying an 'absolute' representation is stronger than the assumption underlying a 'relative' representation. In the studies described in Chapters 4 and 5, the absolute approach presumed that readers would have an overview of the whole text after they prepared it, and that they realized this whole representation prosodically and accordingly. This assumption can only hold when readers prepare a text thoroughly; in other circumstances, it would be an implausible assumption. The relative approach presumed that, even if the readers prepared the text, as speakers they would realize their mental representations of the text prosodically incrementally: segment by segment, the speakers would adjust their prosodic realizations for hierarchy. The results concerning the relation between hierarchy and prosody were found to be compatible with both kinds of representations in the studies described in this dissertation.

Given the fact, as shown in this dissertation, that prepared speakers are able to signal, using pauses and pitch range, the hierarchical levels of text structure, the nuclearity of segments, and various types of rhetorical relations between segments, the relevance for psycholinguistics, and especially for text processing, is that these text-structural features have psychological reality: their reflection in prosody shows that they must be part of speakers' mental representations of the text.

References

- Ayers, G. (1992). Discourse functions of pitch in spontaneous and read speech. Paper presented at the Linguistic Society of America Meeting, Philadelphia (unpublished abstract).
- Bateman, J. A., & Rondhuis, K. J. (1997). Coherence relations: Towards a general specification. *Discourse Processes*, 24, 3-49.
- Batliner, A., Buckow, J. Huber, R., Warncke, V., Nöth, E. Niemann, H. (2001). Boiling down prosody for the classification of boundaries and accents in German and English. *Paper presented at the Eurospeech Conference, Aalborg, Danmark*, 2781-2784.
- Blaauw, E. (1995). *On the perceptual classification of spontaneous and read speech*. Dissertation. University of Utrecht.
- Bond, S. J., & Hayes, J.R. (1984). Cues people use to paragraph text. *Research in the teaching of English*, 18(2), 147-167.
- Brown, G., Currie, K., & Kenworthy, J. (1980). *Questions of intonation*. London: Croom Helm.
- Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.
- Brubaker, R. (1972). Rate and pause characteristics of oral reading. *Journal of psychological research*, 1, 141-147.
- Bruce, G. (1982). Textual aspects of Prosody in Swedish. *Phonetica*, 39, 274-287.
- Carletta, J. (1995). Assessing agreement on classification task: the kappa statistic. *Computational Linguistics*, 22, 249-254.
- Carletta, J., Isard, S., Doherty-Sneddon, G., Isard, A., Kowtko, J., & Anderson, A. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23, 30-31.
- Carletta, J., Isard, A., Isard, S., Kowtko, J., Newlands, A., Doherty-Sneddon, G., Anderson, A. (1995). Dialogue structure coding and its uses in the map task. *Computational Linguistics*, 28(9), 1-44.
- Carlson, L., Marcu, D., & Okurowski, M. (2003). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In R. S. J. van Kuppevelt (Ed.), *Current directions in discourse and dialogue*. (pp. 85-112): Kluwer Academic Publisher.
- Caspers, J. (1994). *Pitch movements under time pressure: effects of speech rate on the melodic marking of accents and boundaries in Dutch*. Dissertation. The Hague: Holland Academic Graphics.
- Caspers, J. (2000). Pitch accents, boundary tones and turn-taking in Dutch Map Task dialogues. *Paper presented at the 6th International Conference on Spoken Language Processing, Beijing, China*, 565-568.
- Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20, 37-46.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213-220.

References

- Condon, S., & Cech, C. (1995). Problems for reliable discourse coding systems. *Paper presented at the Spring Symposium Series of the American Association for Artificial Intelligence, Stanford University.*
- Cooper, W. E., & Sorensen, J.M. (1977). Fundamental frequency contours at syntactic boundaries. *Journal of Acoustic Society of America*, 62(3), 683-692.
- Cooper, W. E., & Sorensen, J. M. (1981). *Fundamental frequency in sentence production*. New York: Springer Verlag.
- Di Cristo, A., Auran, C., Bertrand, R., Chanet, C., & Portes, C. (2002). An integrative approach to the relations of prosody to discourse: towards a multilinear representation of an interface network. *Paper at Laboratoire Parole et Langage, 27March, France.*
- Donzel, M. van (1999). *Prosodic aspects of information structure in discourse*. Dissertation. University of Amsterdam.
- Donzel, M. van, & Koopmans, F.J. (1995). Evaluation of discourse structure on the basis of written vs. spoken material. *Paper presented at the International Congress of Phonetic Sciences, Stockholm, Sweden*, 258-261.
- Flammia, G. (1998). *Discourse segmentation of spoken dialogue: an empirical approach*. Dissertation. Massachusetts Institute of Technology.
- Flammia, G., & Zue, V. (1995). Empirical evidence of human performance and agreement in parsing discours constituents in spoken dialogue. *Paper presented at the Eurospeech Conference, Madrid, Spain*, 1965- 1968.
- Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411-458.
- Geluykens, R., & Swerts, M. (1994). Prosodic cues to discourse boundaries in experimental dialogues. *Speech Communication*, 15, 69-77.
- Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics*, 21, 501-529.
- Grosz, B., & Hirschberg, J. (1992). Some intonational characteristics of discourse structure. *Paper presented at the International Conference on Spoken Language Processing, Banff, Canada*, 429-432.
- Grosz, B., & Sidner, C. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12, 243-281.
- Grosz, B. (1974). The Structure of Task Oriented Dialogs, *IEEE Symposium on Speech Recognition: Contributed Papers*. Pittsburgh: PA, 250-253.
- Gussenhoven, C., Repp, B., Rietveld, A., Rump, H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of Acoustic Society of America*, 102, 3009-3022.
- Gussenhoven, C., & Rietveld, T. (2000). The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech*, 43, 183-203.
- Haan, J., Heuven, V. van, Pacilly, J., & Bezooijen, R. van (1997). An anatomy of Dutch question intonation. *Linguistics in the Netherlands*, 14, 97-108.
- Harley, T. A. (2001). *The psychology of language: from data to theory*. New York: Taylor & Francis.

- Hart, J. 't, Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: an experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.
- Hearst, M. (1997). TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1), 33-64.
- Herwijnen, O. van, & Terken, J. (2001). Do speakers realize the prosodic structure they say to do? *Paper presented at the Eurospeech Conference, Aalborg, Denmark*, 959-962.
- Hirschberg, J., & Grosz, B. (1992). Intonational features of local and global discourse structure, *Proceedings of the speech and natural language workshop*. Harriman NY, DARPA, Morgan Kaufmann, 441-446.
- Hirschberg, J., & Nakatani, C. (1996). A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of the 34th annual meeting Association for Computational Linguistics, Santa Cruz*, 286-293.
- Johns-Lewis, C. (1986). Prosodic differentiation of discourse modes. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 199-219). San Diego: College-Hill Press.
- Koopmans, F. J., & Donzel, M. van (1996). Discourse structure and its influence on local speech rate. *Paper presented at the IFA Proceedings*, 1-11.
- Ladd, R. (1988). Declination 'reset' and the hierarchical organization of utterances. *Journal of Acoustic Society of America*, 84(2), 530-544.
- Ladd, R. (1990). Metrical representation of pitch register. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology: Between the grammar and physics of speech* (pp. 35-57). Cambridge: Cambridge University Press.
- Ladd, R. (1993). On the theoretical status of the baseline in modeling intonation. *Language and Speech*, 36, 435-451.
- Lehiste, I. (1975). The phonetic structure of paragraphs. In A. Cohen & S. Nootboom (Eds.), *Structure and process in speech perception* (pp. 195-203). Berlin: Springer.
- Lehiste, I. (1979). Perception of sentence and paragraph boundaries. In B. Lindblom & S. Ohman (Eds.), *Frontiers of speech communication research* (pp. 191-201). London: Academic Press.
- Levelt, W. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.
- Lieberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & Oehrle (Eds.), *Language Sound Structure* (pp. 157-233). Cambridge MA: MIT Press.
- Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., & Miller, M. (1985). Measurements of the sentence intonation of read and spontaneous speech in American English. *Journal of Acoustical Society of America*, 77, 649-657.
- Mann, B., & Thompson, S. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8, 243-281.
- Marcu, D. (1999). Discourse trees are good indicators of importance in text. In I. Mani, & M. Maybury (Eds.), *Advances in automatic text summarization* (pp. 123-136): MIT Press.
- Marcu, D. (2000). *Theory and practice of discourse parsing and summarization*: MIT Press.
- Marcu, D., & Echihiabi, A. (2002). An unsupervised approach to recognizing discourse relations. *Paper presented at the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia*.

References

- Marcu, D., Romera, M., & Amorrortu, E. (1999). Experiments in Constructing a Corpus of Discourse Trees: Problems, Annotation Choices, Issues. *The ACL'99 Workshop on Standards and Tools for Discourse Tagging*, Maryland, 48-57.
- Möhler, G., & Mayer, J. (2001). A discourse model for pitch-range control. *Paper presented at the 4th ISCA workshop on Speech Synthesis, Perthshire, Scotland.*
- Mozziconacci, S. (1998). *Speech variability and emotion: production and perception*. Dissertation. Technische Universiteit Eindhoven.
- Murray, J.D. (1997). Connectives and narrative text: the role of continuity. *Memory & Cognition*, 25(2), 227-236.
- Mushin, I., Stirling, L., Fletcher, J., & Wales, R. (2003). Discourse structure, grounding, and prosody in task-oriented dialogue. *Discourse Processes*, 35(1), 1-31.
- Nakatani, C., Grosz, B., Ahn, D., & Hirschberg, J. (1995). *Instructions for Annotating Discourses* (pp. 21-95). Harvard University: Center for Research in Computing Technology.
- Noordman, L., Dassen, I., Swerts, M., & Terken, J. (1999). Prosodic markers of text structure. In K. van Hoek, & A. Kibrik & L. Noordman (Eds.), *Discourse studies in cognitive linguistics* (pp. 133-145). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Ouden, H. den, Noordman, L., & Terken, J. (2003). De prosodische realiseren van hiërarchische structuur, nucleariteit en retorische relaties in teksten. *Gramma/TTT*, 9(2/3), 185-209.
- Ouden, H. den, & Terken, J. (2001). Measuring pitch range. *Paper presented at the Eurospeech Conference, Aalborg, Denmark*, 91-94.
- Ouden, H. den, Terken, J., Wijk, C. van, & Noordman, L. (accepted). Betrouwbaarheid bij het meten van toonhoogtebereik. *Gramma/TTT*.
- Ouden, H. den, Noordman, L., & Terken, J. (2002). The prosodic realization of organizational features of texts. *Paper presented at the Speech Prosody Conference, Aix-en-Provence, France*, 543-546.
- Ouden, H. den, & Wijk, C. van (2000). Tekstanalyse met de Rhetorical Structure Theory: een onderzoek naar betrouwbaarheid van structuurtoekenningen. In R. Neutelings & N. Ummelen & F. Maes (Eds.), *Over de grenzen van de taalbeheersing; Onderzoek naar taal, tekst en communicatie* (pp. 148-155). 's-Gravenhage: SDU.
- Ouden, H. den, Wijk, C. van, Terken, J., & Noordman, L. (1998). Reliability of discourse structure annotation. *IPO Annual Progress Report*, 33, 129-138.
- Oviatt, S. L., & Cohen, P.R. (1991). Discourse structure and performance efficiency in interactive and non-interactive spoken modalities. *Computer Speech and Technology*, 5, 297-326.
- Passoneau, R., & Litman, D. (1993). Intention-based segmentation: human reliability and correlation with linguistic cues. *Paper presented at the 31st Annual Meeting of the Association for Computational Linguistics*, 148-155.
- Passoneau, R., & Litman, D. (1997). Discourse segmentation by human and automated means. *Computational Linguistics*, 23, 103-139.
- Pierrehumbert, J. (1979). *The perception of fundamental frequency*. New York: Garland Press.
- Pierrehumbert, J. (1980). The phonetics and phonology of English intonation. *Journal of the Acoustical Society of America*, 66, 363-369.

- Pierrehumbert, J. (1981). Synthesizing intonation. *Journal of the Acoustical Society of America*, 70, 985-995.
- Popping, R. (1996). AGREE 6 for nominal scale agreement. *Groningen: iecProGAMMA*, 12-13.
- Portes, C., & Di Cristo, A. (2003). Pitch range in spontaneous speech: semi-automatic approach versus subjective judgement. *Paper presented at the 15th International Conference on Phonetic Sciences, Barcelona, Spain*.
- Portes, C., Rami, E., Auran, C., & Di Cristo, A. (2002). Prosody and discourse: a multi-linear analysis. *Paper presented at the Speech Prosody Conference, Aix-en-Provence, France*, 579-582.
- Rietveld, T., & Hout, R. van (1993). *Statistical techniques for the study of language and language behavior*. Berlin: Mouton de Gruyter.
- Rotondo, J. A. (1984). Clustering analysis of subjective partitions of text. *Discourse Processes*, 7, 69-88.
- Sanderman, A. (1996). *Prosodic phrasing*. Dissertation. Technische Universiteit Eindhoven.
- Sanders, T., & Noordman, L. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29, 37-60.
- Sanders, T., & Wijk, C. van (1996). PISA, a procedure for analyzing the structure of explanatory texts. *Text*, 16, 91-132.
- Sanders, T. (1992). *Discourse structure and coherence. Aspects of a cognitive theory of discourse representation*. Dissertation. Tilburg University.
- Sanders, T., Spooren, W., & Noordman, L. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, 15, 1-35.
- Schilperoord, J. (1996). *It's about time*. Dissertation. Utrecht University.
- Shimojima, A., Katagiri, Y., Koiso, H., & Swerts, M. (1999). An experimental study on the informational and grounding functions of prosodic features of Japanese echoic responses. *Proceedings of the ESCA Workshop on Dialogue and prosody, Veldhoven, The Netherlands, September 1999*, 187-192.
- Siegel, S., & Castellan, N. (1988). *Nonparametric statistics for the behavioral sciences. Second edition*. New York: McGraw-Hill.
- Silverman, K. (1987). *The structure and processing of fundamental frequency contours*. Dissertation. Cambridge University.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBi: A Standard for Labeling English Prosody. *Proceedings of the Second International Conference on Spoken Language Processing*, 270-286.
- Singer, M. (1990). *Psychology of language*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Sluijter, A., & Terken, J. (1993). Beyond sentence prosody: paragraph intonation in Dutch. *Phonetica*, 50, 180-188.
- Smith, C., & Hogan, L. (2001). Variation in final lengthening as a function of topic structure. *Paper presented at the Eurospeech Conference, Aalborg, Denmark*, 955-958.
- Sorensen, J., & Cooper, W. (1980). Syntactic coding of fundamental frequency in speech production. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 399-440). Hillsdale: Lawrence Erlbaum.

References

- Sweetser, E. (1990). *From etymology to pragmatics*. Cambridge: Cambridge University Press.
- Swerts, M. (1993). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30, 485-496.
- Swerts, M. (1995). *Prosodic features of discourse units*. Dissertation. Technische Universiteit Eindhoven.
- Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America*, 101, 514-521.
- Swerts, M., Bouwhuis, D., & Collier, R. (1996). Melodic cues to the perceived 'finality' of utterances. *Journal of Acoustic Society of America*, 96(4), 2064-2075.
- Swerts, M., & Collier, R. (1992). On the controlled elicitation of spontaneous speech. *Speech Communication*, 11, 463-468.
- Swerts, M., & Geluykens, R. (1993). The prosody of information units in spontaneous monologue. *Phonetica*, 50, 189-196.
- Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, 37(1), 21-43.
- Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. *Proceedings of the 4th International Conference on Spoken Language Processing*, 1033-1036.
- Terken, J. (1984). The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech*, 27, 269-289.
- Terken, J. (1993). Synthesizing natural sounding intonation for Dutch: rules and perceptual evaluation. *Computer Speech and Language*, 7, 27-48.
- Thorndyke, P. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9, 77-110.
- Thorsen, N. (1979). Interpreting raw fundamental frequency tracings of Danish. *Phonetica*, 36, 57-78.
- Thorsen, N. (1985). Intonation and text in Standard Danish. *Journal of Acoustic Society of America*, 77, 1205-1216.
- Wichmann, A. (1991). *Beginnings, middles, ends: intonation in text and discourse*. Dissertation. Lancaster University.
- Wichmann, A., House, J., & Rietveld, T. (1997). Peak displacement and topic structure. *Paper presented at the ESCA Workshop on Intonation: Theory, models and applications, Athens, Greece*, 329-332.
- Wijk, C. van (2000). *Toetsende statistiek: basistechnieken. Een praktijkgerichte inleiding voor onderzoekers van taal, gedrag en communicatie*. Bussum: Coutinho.

Appendix A Original Dutch text of the sample text used in Chapter 2

1 Nederland loopt weer eens gelijk op met de Verenigde Staten van Amerika
2 althans, het Nederland, zoals VVD-leider Bolkestein zich dat droomt
3 de afgelopen week had hij het in het televisieprogramma Netwerk over het failliet van het
4 minderhedenbeleid
5 hij mocht in dat programma verschijnen
6 omdat hij net een boekje heeft uitgebracht met interviews met succesvolle moslims
7 dit is een beproefde strategie van Bolk
8 kruip tegen moslims aan
9 laat zien hoe geweldig je ze vindt
10 en heb het vervolgens alleen nog maar over hoe Nederland zijn minderheden veel strenger moet
11 bejegenen, teneinde ware integratie te bewerkstelligen
12 en niemand kan 'm ergens van beschuldigen
13 want hij heeft tenslotte de Marokkaanse Oussama Cherribi de VVD en de Tweede Kamer binnengehaald
14 hij schreef immers een boek met de Algerijnse hoogleraar Islam, Mohammed Arkoun
15 en nu heeft hij weer een boek gepubliceerd over succesvolle moslims
16 en al deze boeken en acties grijpt hij aan om zijn gelijk te bewijzen
17 normale migranten lukt het zonder steun van de overheid
18 die hebben dat helemaal niet nodig
19 zo pleitte hij in Netwerk ook voor het ophouden met speciale aandacht en overheidssteun voor minderheden
20 die steun moet voor alle achterstandsgroepen gelden
21 want hulp helpt niet
22 wie echt wil, komt er toch wel, op eigen kracht
23 maar meneer Bolkestein, waarom dan niet meteen elke vorm van overheidssteun afschaffen
24 weggegooid geld, toch
25 in de Verenigde Staten heeft Oprah Winfrey via een omgekeerde actie helaas hetzelfde resultaat bereikt
26 als Bolkestein hier
27 zij ging juist bewijzen dat zwarte Amerikanen met een achterstandspositie met wat extra hulp wel degelijk
28 aan een menswaardig bestaan te helpen zijn
29 ze smeet er één miljoen dollar tegenaan
30 en zette een enorme organisatie, vol met pedagogen, psychologen en andere deskundigen, op om zeven
31 zwarte gezinnen boven Jan te helpen
32 het plan mislukte jammerlijk
33 Oprah hield ermee op
34 toen een van de vrouwen die bol stond van de schulden, weigerde om haar mobiele telefoon weg te doen
35 nu zijn de echte hulporganisaties woedend, zoveel geld voor zo weinig mensen
36 terwijl de rest van Amerika zegt: "zie je wel, die mensen zijn niet te helpen"
37 en weer gretig op Oprah zelf wijzen
38 kijk, haar is het wel gelukt, en zonder enige hulp, toch
39 Bolkestein en Oprah, tegengestelde acties, met hetzelfde resultaat en ogenschijnlijk dezelfde conclusie
40 steun helpt niet
41 motivatie, daar draait het om
42 wanneer komt de een of andere slimmerik op het idee dat het heel misschien wel en-en is, hulp en
43 motivatie, en geen of-of

Bron: *Column 'De toestand van de media' (NCRV) door Joan de Windt (journaliste) in het radioprogramma 'Tulpen en olijven', 23 mei 1997*

Appendix B Original Dutch text of the sample text used in Chapter 4

1 Clinton is vanochtend z'n eerste Romeinse dag begonnen alsof ie thuis was, dus met een partijtje joggen
2 naast hem pufte de Amerikaanse ambassadeur in Rome
3 die hem onder het rennen de schoonheden van de eeuwige stad uitlegde
4 en ietsje verder renden de veiligheidsagenten met een pistool in hun short
5 na het omkleden maakte Clinton zijn opwachting bij president Scalfaro
6 met wie hij sprak over democratie en mensenrechten
7 een stuk minder gemakkelijk was daarna het gesprek met de paus
8 dat ging vooral over de VN-conferentie in september in Cairo over bevolking en ontwikkeling
9 in het voorbereidende document voor die conferentie kiest de VN partij voor voorbehoedmiddelen en
10 abortus als middelen om de bevolkingsexplosie in de Derde Wereld terug te brengen
11 dat document heeft de steun van Clinton
12 en 't heeft de woede opgewekt van de paus
13 al maanden voert Johannes Paulus de Tweede een kruistocht tegen deze aanpak van het
14 bevolkingsprobleem
15 die volgens hem neerkomt op moord en op het vernietigen van het gezin
16 president Clinton verzekerde de paus dat ook voor de Amerikanen het gezin centraal staat
17 en hij liet merken dat hij een document van de Amerikaanse katholieke kerk heel serieus neemt
18 daarin staat dat de katholieken in de Verenigde Staten het standpunt van hun president over de
19 bevolkingsproblematiek nooit zullen kunnen delen
20 in de praktijk zijn Clinton en de paus nauwelijks tot elkaar gekomen
21 op een persconferentie heeft Clinton dat zojuist ook toegegeven
22 maar hij zei erbij dat er wel overeenstemming is over de noodzaak tot een duurzame ontwikkeling van de
23 Derde Wereld
24 de persconferentie werd gehouden na afloop van een gesprek tussen Clinton en premier Berlusconi
25 de Italiaanse premier zei dat er van fascisme in zijn regering geen sprake is
26 het is een vals probleem, zei hij
27 uit een opiniepeiling blijkt dat slechts 0 komma 4 procent van de Italianen heimwee heeft naar het
28 fascisme
29 bovendien, zei hij, zijn al mijn ministers democratisch
30 en vinden zij allemaal dat het totalitarisme moet worden bestreden

Bron: *Radiojournaal (AVRO) door Jan van der Putten (verslaggever), 4 April 1994*

Appendix C Original Dutch text of the sample text used in Chapter 5

- 1 De volkstelling in China is met vijf dagen verlengd.
- 2 Die had vrijdag moeten eindigen.
- 3 Maar miljoenen mensen hebben de enquêteurs ontlopen
- 4 of deden de deur niet open.
- 5 Die boycot is vooral bedoeld om illegale kinderen of woonplaatsen geheim te houden.
- 6 Tot de verlenging werd vrijdag besloten op een spoedvergadering van de staatsraad, het Chinese kabinet.
- 7 Een functionaris van de Pekinese volkstellingscommissie zei dat ze hadden gemerkt dat het heel lastig was mensen met drukke bezigheden overdag of 's avonds thuis te treffen, maar dat natuurlijk veel anderen hen opzettelijk ontliepen.
- 8 Minstens tachtig miljoen boeren wonen illegaal in de steden.
- 9 Ze laten zich tellen in hun plaats van herkomst,
- 10 of ze worden helemaal niet geteld.
- 11 Hoewel hen officieel verzekerd is dat de volkstelling buiten de politie omgaat,
- 12 zijn veel mensen bang voor represailles als blijkt dat ze geen verblijfsvergunning hebben.
- 13 Ook veel echtparen met meer dan één kind boycotten de census,
- 14 omdat ze bang zijn dat de geboortebeperkingsdienst er achter komt.
- 15 De demografen van deze dienst geven nu toe dat de meeste mensen zich niet hebben gehouden aan de één-kind-politiek.
- 16 In de provincie Shanxi is dezer dagen zelfs een gezin met tien kinderen gesignaleerd.
- 17 Het tellen van de daklozen levert volgens de autoriteiten weinig problemen op:
- 18 veel zouden het er niet zijn,
- 19 en de meesten zouden wel een dak hebben in een andere provincie.
- 20 Daar zouden ze dan geteld worden.
- 21 Hoe dat moet gebeuren, is echter onduidelijk.
- 22 Anderen voeren een inbreuk op hun privacy aan.
- 23 Deze groep is vooral te vinden onder de willekeurig geselecteerde 10 procent van de bevolking die 49 gedetailleerde vragen krijgt voorgelegd.
- 24 De overige 90 procent hoeft alleen negentien algemene vragen te beantwoorden.
- 25 Advertenties roepen hen die nog niet geteld zijn op zich te melden bij de volkstellingsdienst.
- 26 Het lijkt niet waarschijnlijk dat degenen die de enquêteurs tot nu toe zijn ontlopen op die oproep zullen ingaan.
- 27 Het zal de accuratesse van deze vijfde census in de 51-jarige geschiedenis van de Volksrepubliek niet ten goede komen.

Bron: *de Volkskrant*, 3 November 2000

Appendix D Instruction pretest: causality test

Wat is causaal?

Een causale relatie tussen twee zinnen betekent dat er een oorzaak-gevolg relatie of een probleem-oplossing relatie bestaat tussen de zinnen.

Voorbeeld:

- (1) oorzaak-gevolg: Jan is ziek, (dus) hij is niet op school.

Je zou deze zin kunnen herschrijven als: 'de oorzaak dat Jan niet op school is, is dat hij ziek is.'

- (2) probleem-oplossing: De auto is stuk, (dus) ik breng 'm naar de garage.

Je zou deze zin kunnen herschrijven als: 'de oplossing voor het probleem dat de auto stuk is, is dat ik 'm naar de garage breng.'

Wat is additief?

Een additieve relatie tussen twee zinnen betekent dat er een opsommingsrelatie of een lijst-achtige relatie tussen de zinnen bestaat.

Voorbeeld:

In Turkije is veel zon en cultuur. Aan natuur is geen gebrek. Turkse gastvrijheid voelt aan als een warm bad.

Je zou deze zin kunnen herschrijven als: 'Behalve de zon biedt Turkije veel cultuur. Ook aan natuur is geen gebrek. Daar komt nog bij dat de gastvrijheid in Turkije als een warm bad aanvoelt.'

Voorbeeld vraag causaliteitstest:

Instructie:

Geef aan of je vindt dat de vetgedrukte zin een *causale* relatie of een *additieve* relatie onderhoudt met de voorafgaande zin. Maak het bijbehorende rondje zwart.

Tekst:

Sinds de overstromingen van 1997 is Rijkswaterstaat druk geweest met het ontwikkelen van oplossingen voor wateroverlast. Er worden natuurvriendelijke oevers aangelegd. Stuwen in het traject hebben een remmend effect op water. **Het gebruik van mobiele noodpompen moet zorgen voor een betere waterafvoer in tijden van extreem veel neerslag.** De dijken langs het traject worden opgehoogd tot Delta-niveau. Op deze manier hoopt men dat Nederland behoed wordt voor overstromingen van rivieren en sloten.

causaal

0

additief

0

Appendix E Instruction pretest: plausibility test

Experiment over causale en niet-causale relaties:

Voorbeeld vraag plausibiliteitstest:

Instructie:

Geef aan of je vindt dat de vetgedrukte zin natuurlijk volgt op de voorafgaande zin.

Maak het bijbehorende rondje zwart.

Tekst:

Het aantal evenementen is met de komst van de Floriade weer gestegen. De opbrengst van deze tentoonstellingen is wisselend. De Huishoudbeurs is erg stabiel gebleken. De Vakantiebeurs groeit ieder jaar nog met 5 procent. **De Floriade verwacht een miljoenenverlies.** Ook de RAI-beurzen verwachten geen winst. Kleinere exposities doen het lokaal prima, maar tellen nauwelijks mee op landelijk niveau.

heel natuurlijk	natuurlijk	neutraal	onnatuurlijk	heel onnatuurlijk
0	0	0	0	0

Experiment over semantische en pragmatische relaties:

Voorbeeld vraag plausibiliteitstest:

Instructie:

Geef aan of je vindt dat de vetgedrukte zin natuurlijk volgt op de voorafgaande zin.

Maak het bijbehorende rondje zwart.

Context:

Je vertelt een medestudent over je huisgenoot Walter. Je medestudent kent Walter ook. Walter heeft je vanochtend verteld dat hij ziek is. Je legt aan je medestudent uit waarom Walter er vandaag niet is.

Tekst:

We moeten het vandaag zonder Walter stellen. Hij komt de hele dag niet naar school.

Hij is ziek. Hij heeft de hele avond liggen hoesten.

heel natuurlijk	natuurlijk	neutraal	onnatuurlijk	heel onnatuurlijk
0	0	0	0	0

Appendix F Instruction pretest: semanticity test

Wij zouden graag uw oordeel krijgen over de onderstaande teksten. Alle teksten worden voorafgegaan door een context. Deze context geeft de situatie aan waarin u zich moet inleven wanneer u de teksten beoordeelt. De teksten bevatten een zinspaar van twee semantisch gerelateerde zinnen of van twee pragmatisch gerelateerde zinnen. In alle gevallen gaat het om causale relaties. De vraag is of u de relaties als semantisch of als pragmatisch beoordeelt.

Het gaat om de relatie tussen de cursief gedrukte zin(nen) en de vetgedrukte zin. Onder iedere tekst staat de onderstaande tabel waarin u elektronisch, door middel van een kruisje in de onderste rij, uw oordeel kunt geven. De betekenis van de schaal is: 2= sterk, 1= zwak en 0= onduidelijk. Het ingevulde formulier kunt u via e-mail terugsturen.

Semantisch	2	1	0	1	2	Pragmatisch
------------	---	---	---	---	---	-------------

Een voorbeeld van een semantische relatie is:

(1) Jan is niet op zijn werk omdat hij ziek is.

Wanneer twee segmenten semantisch met elkaar verbonden zijn, hangen de segmenten met elkaar omdat de beschreven gebeurtenissen in de werkelijkheid met elkaar samenhangen. In (1) baseert de schrijver zich op het feit dat hij weet dat Jan ziek is, Jan heeft hem bijvoorbeeld gebeld om dat te zeggen. Jan is daadwerkelijk ziek.

Een voorbeeld van een pragmatische relatie is:

(2) Jan is ziek omdat hij niet op zijn werk is.

Wanneer twee segmenten pragmatisch met elkaar verbonden zijn hangen de segmenten met elkaar samen op basis van een conclusie die de spreker of schrijver trekt. In (2) trekt de spreker een conclusie uit het feit dat Jan niet op zijn werk is. In (2) redeneert de spreker als volgt: Iemand die ziek is komt niet naar zijn werk. Jan is niet op zijn werk, dus Jan moet wel ziek zijn.

Het verschil tussen (1) en (2) zit in het feit dat het in (1) bekend is dat Jan ziek is, terwijl de spreker in (2) weliswaar denkt dat Jan ziek is, maar er kan ook een andere reden zijn voor Jans afwezigheid.

Van de onderstaande teksten willen wij dus van u weten of u de causale relatie tussen de cursieve zin(nen) en de vetgedrukte zin als semantisch dan wel als pragmatisch beoordeelt.

Ik dank u hartelijk voor uw medewerking.

Appendix G Texts used in the experiment on causal and non-causal relations

Causal relations

1

Viola was niet erg handig. Een maand of drie geleden had ze een tijd lang met dof haar gelopen omdat ze zo nodig had willen besparen op kapperskosten. Nu was het weer gezond, maar ze wilde nog steeds wel een lichtere tint. **Ze liet haar haren blonderen bij de kapper.** Ze zag er jaren jonger uit en voelde zich geweldig. Viola ging de stad in en kocht ook nog wat nieuwe kleren en een paar prachtige laarzen.

2

De deelstaat Mecklenburg-Voor-Pommeren heeft erg veel last van het wassende water. Veel dijken zijn doorweekt en kunnen mogelijk doorbreken. **Het leger wordt ingezet om de dijken te verstevigen.** Op deze manier hoopt Duitsland de schade te beperken. Honderden soldaten hebben zich inmiddels verzameld om de klus te klaren. Zij zijn tijdelijk ondergebracht in barakken in de buurt van het rampgebied.

3

Opnieuw is een inwoner van Veendam omgekomen bij een verkeersongeval. De man stak de straat over en werd aangereden door een vrachtwagen. Inwoners van Veendam zijn al jaren bezig om de verkeersoverlast in de stad terug te dringen. De overlast wordt voornamelijk veroorzaakt door vrachtverkeer dat door het centrum van de stad moet. **De aanleg van een tunnel in het centrum van Veendam zal begin volgend jaar aanvangen.** Dit is besloten in de raadsvergadering van afgelopen dinsdag. Als de tunnel klaar is, zullen voetgangers en fietsers weer veilig het centrum van Veendam kunnen bereiken.

4

Gisterochtend werd Breda opgeschrikt door een hevig onweer vergezeld van zeer zware regenval. Het zicht op de wegen was minimaal. **De opritten naar de snelweg werden tijdelijk afgesloten.** Na een vijftiental minuten stopte de bui even plotseling als hij begonnen was. Ondanks de barre weersomstandigheden zijn er geen ongelukken gebeurd. Het KNMI noemde het noodweer een samenloop van atmosferische storingen.

Non-causal relations

1

Volgende week zou Viola haar ja-woord geven. Vandaag werd een drukke dag met veel voorbereidingen. Viola ging naar de schoonheidsspecialiste voor een behandeling. Bij de nagelstudio werden speciale harsnagels aangebracht. **Ze liet haar haren blonderen bij de kapper.** Die zou ook meteen zorgen voor een haarband met strasssteentjes, want die had Viola niet. Daarna zou haar toekomstige schoonzusje komen om de jurk aan te passen.

2

Bij Defensie is de kans op een veelzijdige baan groot. Naast deelname aan vredesmissies zijn ook maatschappelijke problemen een uitdaging voor de militair. Er wordt een beroep op defensie gedaan bij calamiteiten. Mislukte oogsten worden van het land gehaald door soldaten. **Het leger wordt ingezet om de dijken te verstevigen.** Daarnaast helpt Defensie regelmatig bij opsporingen van vermiste kinderen

3

Verkeer in de regio Groningen zal begin volgend jaar rekening moeten houden met vertragingen en opstoppen door aanleg en onderhoud van de wegen. Tussen Stadskanaal en Veendam wordt een nieuwe regionale weg aangelegd. De afrit van de snelweg tussen Veendam en de Duitse grens zal van nieuw asfalt worden voorzien. **De aanleg van een tunnel in het centrum van Veendam zal begin volgend jaar aanvangen.** Met deze werkzaamheden nadert de voltooiing van het project "Groningen beter bereikbaar", dat in 1998 gestart is.

4

Bij de onderhoudswerkzaamheden van de snelweg luisterde de planning erg nauw. Daarom werden er eerst rijdende afzettingen geplaatst. Daarna moesten er mobiele verkeerslichten gezet worden. **De opritten naar de snelweg werden tijdelijk afgesloten.** Op sommige trajecten was maar één rijbaan beschikbaar. Tijdens de spits werden automobilisten omgeleid via een andere route.

5

Klimaatveranderingen hebben grote gevolgen voor de wereldbevolking. Door het broeikaseffect ontstaat een stijging van het waterpeil van de aardbol. Juist aan de Aziatische zijde van de aarde stijgt het water opzienbarend. **In China moeten tienduizend mensen geëvacueerd worden.** De regering beraadt zich op manieren om een eventuele evacuatie op korte termijn te realiseren. Daarbij vraagt de Chinese regering hulp van de omliggende landen en de VN.

6

In Italië is een heftig debat aan de gang over de stand van de inflatie. Italianen zijn gewend dat munten helemaal niets waard zijn en geven muntgeld uit als water. **Italië wil dat het muntgeld van 1 en 2 euro vervangen wordt door papiergeld.** Gebruikers van de euro moeten een nieuw 'mentaal ijkpunt' krijgen om de munt op waarde te kunnen schatten. Op deze manier hoopt men de koopkracht intact te houden.

7

Noodgedwongen investeringen kleuren de cijfers van NS rood. In een persconferentie gaven ze toe dat veel van het materieel vervangen moet worden. **Trein en bus worden fors duurder.** Het is nog niet bekend hoe consumentenorganisaties reageren op deze actie. De maatregel komt erg slecht uit omdat door verschillende instanties het gebruik van openbaar vervoer gestimuleerd zou gaan worden.

8

De politie heeft een grote groep studenten aangehouden die bij de ontgroening vernielingen aangericht hebben in de binnenstad. Helaas kampt de politie met een ernstig cellentekort. Vijftientwintig studenten zijn ondergebracht in het asielzoekerscentrum. De andere vijftientwintig zijn verdeeld over politiebureaus in West-Brabant. Politie West- en Midden-Brabant heeft al diverse keren haar ongenoegen geuit over deze gang van zaken. Het probleem van het cellentekort staat hoog op de agenda van de verantwoordelijke instanties.

5

Het jaar 2002 gaat de geschiedenis in als het jaar van de wateroverlast. Wereldwijd zijn gebieden getroffen door extreme regenval en overstromingen. Duitsland heeft inmiddels de waterstroom onder controle. In Peru wordt met man en macht geprobeerd het water binnen de oevers te houden. **In China moeten tienduizend mensen geëvacueerd worden.** Ook in ons land is het water een zorg: boeren krijgen de gewassen niet op tijd de grond uit.

6

Uit alle landen waar de Euro zijn intrede deed, zijn reacties gekomen. In Zweden werden de 1- en 2-centsmunten snel na de invoering al verbannen. Nederland pleit voor afschaffing van alle munten onder de 10 eurocent. **Italië wil dat het muntgeld van 1 en 2 euro vervangen wordt door papiergeld.** België is tevreden omdat het Belgische geld al muntstukken had die erg veel leken op de euromunten.

7

De inflatie is voelbaar in de gehele toeristische sector, zowel in vervoer als verblijf. De hotels verdubbelen bijna de prijzen. Vliegmaatschappijen schroeven de prijzen behoorlijk op. **Trein en bus worden fors duurder.** Appartementen zijn qua prijs niet meer vergelijkbaar met een aantal jaren geleden. Dit is nog maar een greep uit de grote berg aan klachten, die de ANVR de afgelopen drie maanden heeft ontvangen.

8

Studenten die nog geen permanente huisvesting hebben gevonden, kunnen tijdelijk een beroep doen op leegstaande locaties. Ongeveer 15 jongeren mochten naar het oude klooster aan de Beverweg. Een kleine groep kon in een voormalig kraakpand terecht. **Vijftientwintig studenten zijn ondergebracht in het asielzoekerscentrum.** Er zijn zelfs studenten die gewoon bij hun ouders blijven wonen, ook al kost het ze drie uur reistijd per dag. Mindere studieresultaten nemen zij op de koop toe.

9

Maurice hield van dansen en ging regelmatig naar een dancefestival. Op diverse plaatsen in Nederland worden 24 uurs-megafestivals georganiseerd. Vermoeidheid is de grootste vijand. **Maurice slikte meestal twee peppilletjes.** Hij bleef met gemak op de been en hij had het nog leuker ook. Hij was een kleinverbruiker in zijn omgeving. De meesten namen vier pilletjes of meer. Ook een lijntje coke wil wel helpen om de nacht door te komen.

10

Tegenwoordig worden jongeren steeds langer en alles groeit mee. Mensen met een schoenmaat groter dan 46 hebben moeite passend schoeisel te vinden. **In Rotterdam is een schoenwinkel voor grote maten.** De collectie bestaat uit hedendaagse schoenmode. Zelfs de gerenommeerde merken zijn verkrijgbaar in grotere maten. Er is een ruime keuze en de klant kan kiezen uit diverse modellen en kleuren.

11

Een fietser uit Oost-Souburg is gisterochtend gewond geraakt aan zijn knie toen hij in Middelburg werd aangereden door een automobilist. De automobilist kwam vanaf de Koestraat en verleende op de kruising met de Vlissingsestraat geen voorrang aan een fietser die van rechts kwam. **De fiets werd geheel vernield.** De jongen is per ambulance naar het ziekenhuis vervoerd. De verwondingen vielen dusdanig mee, dat hij na behandeling weer naar huis mocht keren.

12

Harry bekeek de deur eens goed. De deur klemde vreselijk. Daarbij maakte de deur een schurend geluid dat door merg en been ging. **Hij schaafde een dun laagje van de deur af.** Juist dat kleine beetje extra ruimte was genoeg om het probleem te verhelpen. Openen en sluiten ging weer soepel en ook het schurende geluid was weg.

9

De meeste jongeren zijn niet heel erg open over hun gebruik van verdovende middelen, maar enkelen vertelden ons wat ze op een dance-avond innamen. Robert, Paul en Boy rookten regelmatig een joint. Kim en Kelly hielden het op herbals. **Maurice slikte meestal twee peppilletjes.** Johan en Sven waren al een tijdje geleden over gegaan op speed. Een hele grote groep partygangers vond het feesten leuker onder invloed van alcohol.

10

Nederlanders worden steeds langer. De middenstand springt daar perfect op in. In Amsterdam is een speciale winkel voor lange mensen. King-size bedden zijn te koop in Loosdrecht. **In Rotterdam is een schoenwinkel voor grote maten.** In Breda kun je terecht voor allerlei meubilair van enorme afmetingen. Steeds meer winkeliers beseffen dat de gemiddelde Nederlander niet meer bestaat en passen hun assortiment aan.

11

Voor een project van de Kunstacademie zijn diverse spectaculaire acties ondernomen. Fietsfabrikant Batavus sponsorde het project en stelde geld en een fiets ter beschikking. De enige voorwaarde voor deelname was dat er een fiets in het kunstwerk zou terugkomen. Voor het project werd een auto in brand gestoken. De zitting van een schommel werd verbogen. **De fiets werd geheel vernield.** Tezamen werden de stukken schroot opgehangen in de schommel. Het kunstwerk heet 'Ravage'.

12

In het programma 'Eigen huis en tuin' liet Nico zien hoe je deuren moet onderhouden. Eerst haalde hij de deur uit de scharnieren. Toen zette hij de deur in de garage. **Hij schaafde een dun laagje van de deur af.** Vervolgens werd de deur met ammoniak bewerkt en in de grondverf gezet. Het aflakken deed Nico in twee keer.

13

In de binnenstad moet een groot aantal parkeerplaatsen verdwijnen. De gemeenteraad geeft de voorkeur aan een wandelpromenade boven parkeergelegenheid in het centrum. **De gemeente laat een parkeergarage aanleggen onder de Voorstraat.** Het zal ruimte bieden aan ruim 250 auto's. Het hele plan mag 3,5 miljoen euro kosten. In de zomer van 2003 moeten de inwoners kunnen genieten van de autovrije binnenstad.

14

Nederland is onlosmakelijk verbonden met water en de problemen daar omheen. De laatste jaren heeft regen al verschillende keren tot wateroverlast geleid. Het duurt te lang voordat de riolen weer vrij zijn. **Het gebruik van mobiele noodpompen moet zorgen voor een betere waterafvoer in tijden van extreem veel neerslag.** Rijkswaterstaat hoopt dat deze oplossing volstaat, maar is wel al bezig met de ontwikkeling van alternatieve middelen, zoals dijkverhoging en stuwen.

15

Hardlopers zijn altijd erg benieuwd naar hun geleverde prestaties. Zij klokken elke kilometer of parkoers en willen graag hun persoonlijke tijden verbeteren. Goede meetapparatuur is daarbij onontbeerlijk. **Nike heeft de Tailwind op de markt gebracht.** Dit apparaatje is een soort chronograaf dat tijdens het rennen tegelijkertijd snelheid, afstand en calorieverbruik aangeeft.

16

De kranten stonden bol van aanrandingen en overvallen. Katja voelde zich kwetsbaar en onveilig op straat. Soms durfde ze niet eens meer naar huis als ze in de stad was. Zij wilde juist graag weerbaar zijn. **Katja schreef zich in voor een cursus karate.** De eerste stap op weg naar een veiliger gevoel. Misschien sliep ze dan 's nachts ook weer wat beter.

13

De gemeenteraad heeft besloten om het centrum van de stad autovrij te gaan maken. Een wandelpromenade heeft de voorkeur. Diverse werkzaamheden staan op stapel. Bestrating wordt vervangen zodat het geschikter is voor voetgangers. Er worden bomen geplant op de markt. **De gemeente laat een parkeergarage aanleggen onder de Voorstraat.** Alle verkeerslichten zullen verdwijnen. Het plan zal 3,5 miljoen euro kosten.

14

Sinds de overstromingen van 1997 is Rijkswaterstaat druk geweest met het ontwikkelen van oplossingen voor wateroverlast. Er worden natuurvriendelijk oevers aangelegd. Stuwen in het traject hebben een remmend effect op water. **Het gebruik van mobiele noodpompen moet zorgen voor een betere waterafvoer in tijden van extreem veel neerslag.** De dijken langs het traject worden opgehoogd tot Delta-niveau. Op deze manier hoopt men dat Nederland behoeft wordt voor overstromingen van rivieren en sloten.

15

Sportprestaties moet je kunnen meten en onderling kunnen bespreken. Diverse fabrikanten spelen op deze behoefte in. De Spinometer is een analoge meter en legt skateprestaties vast. Schmit heeft een multifunctionele snelheidsmeter bedacht, die voor diverse takken van sport is te gebruiken. **Nike heeft de Tailwind op de markt gebracht.** Die meet specifiek de prestaties van hardlopers.

16

Studenten kunnen gebruik maken van de Studenten Sport School om aan hun conditie te werken. Carolien wilde graag gaan skaten. Marieke, Martine en Mechteld deden op maandagavond mee met aerobics. **Katja schreef zich in voor een cursus karate.** Hans en Bart gingen op yoga. De sportkaart kost voor studenten maar €40 per jaar.

Appendix H Texts used in the experiment on semantic and pragmatic relations

Pragmatic relations

1

Context: Samen met een klasgenoot heb je het over jullie klasgenoot Rick. Jullie denken dat Rick gevoelens heeft voor Els, maar jullie weten het niet zeker. Rick en Els hebben jullie nooit iets in die richting verteld.

Tekst:

Rick wordt altijd rood als ik het over Els heb. Hij begint ook telkens te stotteren als hij met haar praat. **Hij is hartstikke verliefd.** Volgens mij is Els niet in hem geïnteresseerd.

2

Context: Je vertelt je partner over je collega Kees. Je partner kent Kees ook en mag hem net als jij erg graag. Jullie weten allebei dat Kees een trotse man is die niet graag toegeeft dat iets hem niet lukt. Kees heeft je in de afgelopen tijd niets verteld over de werkdruk.

Tekst:

Kees moet de laatste tijd wel erg veel werken. Ik heb medelijden met hem. Gisteren, onder die vergadering, zag ik dat hij zat te knikkebollen. **Hij kan het niet meer bolwerken.** Ik zit te denken om wat uren van hem over te nemen.

3

Context: Je praat met een klasgenoot over jullie gemeenschappelijke vriend Piet. Jullie weten dat Piet vandaag een presentatie moet houden. Jullie hebben hem wel in de klas zien zitten, maar konden hem niet vragen of hij tegen de presentatie op zag.

Tekst:

Ik zat een eindje bij Piet vandaan. Ik heb hem niet gesproken. Ik zag dat hij de hele tijd op zijn nagels zat te bijten. **Hij is nerveus.** Ik hoop voor hem dat hij zich goed heeft voorbereid.

Semantic relations

1

Context: Een vriendin vraagt hoe het met Jeroen gaat. Je hebt Jeroen pas nog gesproken en hij heeft toen verteld dat hij heel erg verliefd is op Carla. Je vertelt je vriendin wat je van Jeroen hebt gehoord.

Tekst:

Vorige week heb ik Jeroen nog gezien. Het gaat goed met hem. **Hij is hartstikke verliefd.** Carla en hij hebben nu twee maanden een relatie.

2

Context: Je hebt het met een vriend over werken bij een aannemer. Zodoende komt het gesprek op je broer Koen. Koen heeft je vorige week verteld dat hij elders wil gaan werken omdat het werk bij een aannemer te zwaar voor hem is.

Tekst:

M'n broer, je weet wel Koen, werkt sinds twee maanden voor een aannemer. Hij is nu alweer op zoek naar ander werk. **Hij kan het niet meer bolwerken.** Hij heeft liever iets wat minder zwaar is.

3

Context: Een vriend informeert naar je huisgenoot Alex. De vriend weet dat Alex vandaag zijn rij-examen moet afleggen en is benieuwd hoe dat zal gaan aflopen. Vanmorgen heb je Alex nog gesproken. Hij zei toen dat hij nerveus was en dat hij bang was dat hij daardoor vergissingen zou maken. Je vertelt je vriend wat je van Alex hebt gehoord.

Tekst:

Alex moet zich om twee uur bij de rijsschool melden. Hij is bang dat hij domme fouten zal maken. **Hij is nerveus.** Hij zal vanmiddag wel bellen om te vertellen hoe het is afgelopen.

4

Context: 's Avonds als je thuiskomt van je werk vertel je je partner over je collega Klaas. Je partner kent Klaas ook en weet dat hij een harde en trouwe werker is. Je bent erg op Klaas gesteld en je praat dagelijks met hem. Klaas heeft je niet gezegd dat hij ziek is.

Tekst:

Ik schrok toen ik Klaas vandaag zag. Hij zag asgrauw en hij hoestte verschrikkelijk. **Hij is ziek.** Ik zal hem vanavond eens bellen.

5

Context: Met een klasgenoot heb je het over een andere klasgenoot, Bart. Van Bart is bekend dat hij een bijbaantje heeft. Omdat je samen met de klasgenoot aan Bart wil vragen of hij mee wil betalen aan een cadeau, vragen jullie je af of hij veel te spenderen heeft.

Tekst:

Ik heb geen flauw idee of Bart veel verdient met dat bijbaantje. Hij draagt altijd tweedehands kleding. In de kroeg heb ik hem nog nooit een rondje zien geven. **Hij heeft het niet zo breed.** Volgens mij kunnen we er het beste maar eerlijk naar vragen.

6

Context: Je bent samen met je twee kinderen in het zwembad. Eén van je kinderen, Erik, had eerst niet zo'n zin om te gaan. Je partner komt later in het zwembad en vraagt aan jou of Erik het ook leuk vindt. Je hebt het Erik niet gevraagd en daarom weet je het niet zeker, maar je denkt wel dat hij het naar zijn zin heeft.

Tekst:

Erik had er eerst echt geen zin in. Maar toen we aankwamen rende hij gelijk naar de glijbaan. Nu is hij daar met andere kinderen op dat vlot aan het spelen. **Hij heeft het naar zijn zin.** Ik denk wel dat we de hele middag kunnen blijven.

4

Context: Je vertelt een medestudent over je huisgenoot Walter. Je medestudent kent Walter ook. Walter heeft je vanochtend verteld dat hij ziek is. Je legt aan je medestudent uit waarom Walter er vandaag niet is.

Tekst:

We moeten het vandaag zonder Walter stellen. Hij komt de hele dag niet naar school. **Hij is ziek.** Hij heeft de hele avond liggen hoesten.

5

Context: Met een vriend heb je het over vakantiebestemmingen. Het gesprek komt op je broer. Je broer zit al drie jaar zonder werk. Je weet ook dat je broer daardoor niet veel geld heeft. Je vertelt je vriend over je broer.

Tekst:

Bij ons in de familie gaan ze allemaal naar Frankrijk. M'n broer is de enige die zijn vakantie thuis viert. Hij kan het zich niet veroorloven om ieder jaar op vakantie te gaan. **Hij heeft het niet zo breed.** Toch klaagt hij daar nooit over.

6

Context: Je hebt net aan de telefoon gezeten met je beste vriend, Rens, die op stage is in Spanje. Hij heeft je verteld hoe leuk hij het daar vindt. Je moeder kent Rens ook en wil graag weten hoe het met hem gaat. Je vertelt alles zoals je het van Rens hebt gehoord.

Tekst:

Rens werkt bij een bedrijf in hetzelfde stadje als waar hij woont. Hij zei dat hij over drie maanden terugkomt naar Nederland. Na zijn stage zou hij eigenlijk liever daar blijven wonen. **Hij heeft het naar zijn zin.** Als hij straks terug komt moet hij weer bij zijn ouders gaan wonen.

7

Context: Samen met een collega vraag je je af of Koos, jullie collega van de inkoop, door heeft dat er veel kleine dingen worden gestolen uit het magazijn. Jullie hebben Koos hier nog nooit iets over horen zeggen.

Tekst:

Ik weet niet of Koos zo naïef is als wij denken. Ik heb gezien dat hij een nieuw slot op de deur van het magazijn heeft geplaatst. **Hij vermoedt iets.** Ik denk dat iemand hem iets heeft verteld.

8

Context: Tijdens de afwas vertel je je partner over je collega Wouter. Je partner kent Wouter uit jouw verhalen. Ze weet dat je altijd veel moet lachen met Wouter. Vandaag heeft Wouter niet veel gezegd en je weet niet waarom.

Tekst:

Ik vond er vandaag helemaal niets aan op kantoor. Wouter was niet erg spraakzaam tegen mij en tegen de anderen heeft hij zijn mond zelfs helemaal niet opgedaan. **Hij zit ergens mee in zijn maag.** Ik hoop dat hij morgen weer een beetje de oude is.

9

Context: Vanuit een taxi zie je samen met een vriend jullie gemeenschappelijke vriend Piet uit een café komen. Jullie zijn in het verleden vaak met Piet op stap geweest en weten daardoor dat hij graag en veel drinkt. Dat Piet deze avond in de stad zou zijn, wisten jullie niet.

Tekst:

Is dat Piet die daar naar buiten komt? Hij ziet er niet uit. Moet je kijken, hij loopt helemaal te slingeren. **Hij is hartstikke dronken.** Zou hij een feestje van zijn werk hebben gehad?

7

Context: Samen met vrienden had je een surprise-party georganiseerd voor Davids 30^e verjaardag. Een week geleden heeft David jullie betrapt bij de voorbereiding en sindsdien heeft hij een paar keer gevraagd of jullie iets van plan zijn. Je vertelt nu aan iemand wanneer het feest is en dat het voor David geen verrassing meer is.

Tekst:

Iedereen is uitgenodigd op het feest aanstaande zondag. Maar het is geen verrassing meer voor David. **Hij vermoedt iets.** Het is moeilijk om iets te organiseren zonder dat het uitlekt.

8

Context: Je broer Martin heeft je verteld dat hij ergens mee zit en dat hij daarom bij zijn vriendin, Astrid, langs gaat om te praten. Je legt aan je moeder uit waarom Martin later thuis komt.

Tekst:

Ik moet van Martin doorgeven dat hij vandaag een uurtje later thuis komt. Hij wou na school nog even langs bij Astrid. **Hij zit ergens mee in zijn maag.** Hij zei dat hij wel naar huis zou komen om te eten.

9

Context: Je komt 's avonds samen met je huisgenoot Richard thuis van een feest op de hockeyclub. Een andere huisgenoot vraagt je hoe het feest was. Je bent de hele avond met Richard op het feest geweest en je hebt gezien dat Richard erg dronken is geworden. Je vertelt je huisgenoot over Richard.

Tekst:

Richard ligt nu op de bank te slapen. Hij kwam de trap niet meer op. **Hij is hartstikke dronken.** Hij is totaal niet meer aanspreekbaar.

10

Context: Voor de bouw van een nieuw kantoorpand moet een aannemer worden gecontracteerd. Omdat jij veel contacten onderhoudt met aannemers vraagt je baas naar je mening. Hij heeft gehoord over een aannemer en vraagt aan jou of jij hem geschikt vindt. Je kunt niet met zekerheid zeggen of de aannemer een goede partij is. Je baseert je oordeel op dingen die je hier en daar eens over de man hebt gehoord.

Tekst:

Ik weet niet of hij de meest geschikte man is voor die opdracht. Ik heb gehoord dat hij geen kantoor en geen vast personeel heeft en ook dat hij al een keer failliet is gegaan. **Hij is onbetrouwbaar.** Ik zou hem geen opdracht geven.

11

Context: Een goede vriendin van je, Kelly, is net getrouwd. Samen met een andere vriendin vraag je je af of Kelly graag kinderen zou willen hebben. Jullie kennen Kelly allebei goed, maar van een eventuele kinderwens weten jullie niets af.

Tekst:

Ik heb het haar nooit gevraagd, ze heeft mij ook nooit iets in die richting gezegd. Maar ik weet wel dat ze heel goed met de buurkinderen kan omgaan. En ze heeft ook heel lang een baantje gehad bij de crèche. **Ze vindt kinderen leuk.** Laten we het maar eens afwachten.

12

Context: Je werkt op een afdeling voor kwaliteitsonderzoek. Tijdens een functioneringsgesprek vraagt je chef jou naar de samenwerking met je collega Hans. Bij je chef kaart je aan dat het niet goed gaat met Hans. Je hebt gezien dat Hans fouten heeft gemaakt bij zijn werk. Je weet het niet zeker, maar je denkt dat dit komt omdat hij er niet bij is met zijn gedachten.

Tekst:

Ik heb altijd graag samengewerkt met Hans. Maar deze maand heeft hij al vijf keer een fout niet opgemerkt. **Hij is er met zijn gedachten niet bij.** Misschien moet je eens met hem praten.

10

Context: Van een zakenkennis heb je gehoord dat niemand meer zaken wil doen met Herman. Herman blijkt nooit zijn afspraken na te komen en is daardoor erg onbetrouwbaar. Een collega weet hier niks van en vraagt hoe het zit met Herman. Je vertelt je collega wat je van je zakenkennis hebt gehoord.

Tekst:

Het bedrijf van Herman loopt niet goed. Hij heeft zich nooit aan afspraken gehouden. Er is niemand die nog zaken met hem wil doen. **Hij is onbetrouwbaar.** Hij heeft ook moeite om zijn personeel vast te houden.

11

Context: Je partner, Renate, is parttime bij een buurthuis gaan werken waar activiteiten voor allerlei doelgroepen worden georganiseerd. Je weet dat ze het liefst met de kinderen werkt, omdat ze kinderen erg leuk vindt. Een collega heeft iets opgevangen en vraagt jou naar de nieuwe baan van je partner. Je vertelt hem wat je ervan weet.

Tekst:

Renate werkt bij het buurthuis bij ons in de straat. Er worden daar veel activiteiten georganiseerd voor verschillende doelgroepen. Het leukste aan het werk vindt ze de activiteiten met kinderen. **Ze vindt kinderen leuk.** Ze hoeft alleen de middagen te werken.

12

Context: Je vertelt een vriend over het reilen en zeilen in je eigen reclamebedrijf. Je legt hem uit waarom je compagnon Gerard nog niet is begonnen aan een nieuwe opdracht. Jullie weten allebei dat Gerards vader net is overleden en dat Gerard er daarom niet bij is met zijn gedachten.

Tekst:

Gisteren hebben we een mooie opdracht binnengehaald. Gerard begint pas volgende week aan die klus. **Hij is er met zijn gedachten niet bij.** Hij heeft een week vrij genomen.

Samenvatting

Het onderzoek waarover in dit proefschrift wordt gerapporteerd betreft de relatie tussen tekststructuur en prosodie. Tekststructuur heeft te maken met de wijze waarop teksten opgebouwd zijn. Een tekst is een verzameling van zinnen die met elkaar samenhangen. Deze samenhang kan weergegeven worden in een hiërarchische structuur. Een tekstanalyse die gericht is op het verkrijgen van zo'n hiërarchische structuur verloopt als volgt. Op grond van de tekstrelatie die de tekst domineert, wordt de tekst verdeeld in enkele grote teksteenheden; vervolgens wordt op grond van de tekstrelatie die deze teksteenheid domineert de eenheid opgesplitst, en zo verder, tot uiteindelijk de individuele zinnen zijn bereikt. De relaties tussen de zinnen van de tekst zijn dan volledig weergegeven in een hiërarchisch georganiseerde structuur.

Prosodie heeft betrekking op de suprasegmentele kenmerken van spraak. Dit zijn alle eigenschappen van spraak die boven het niveau van de individuele klanken uitgaan zoals pauzes, spreektempo, intonatie, accentuering en volume. Prosodisch onderzoek heeft zich vooral gericht op de kenmerken in geïsoleerde zinnen. Voor het Nederlands bijvoorbeeld, zijn de mogelijke intonatiepatronen van zinnen gedetailleerd beschreven door 't Hart, Collier en Cohen (1990). Er zijn echter aanwijzingen dat bepaalde prosodische kenmerken het niveau van de individuele zinnen overstijgen. Zo is aangetoond dat de pauzes tussen zinnen binnen een alinea korter zijn dan die tussen alinea's (Silverman, 1987; Swerts, 1997), dat de toonhoogte geleidelijk daalt over zinnen binnen een alinea (Bruce, 1982), en dat de spreesnelheid van zinnen in een tekst hoger is dan in isolement (Sanderman, 1996). Daarom wordt in dit proefschrift heel specifiek gekeken naar prosodie op tekstueel niveau, in het bijzonder naar de relatie tussen prosodie en tekststructuur. De centrale vraag is of, analoog aan hoe tekststructuur in een geschreven tekst door de schrijver wordt aangegeven met behulp van allerlei typografische middelen, tekststructuur in een gesproken tekst wordt aangegeven met prosodische middelen.

Met de vraag naar de relatie tussen tekststructuur en prosodie komen twee betrekkelijk van elkaar gescheiden onderzoeksgebieden samen. Om die bij elkaar te kunnen brengen waren enkele voorbereidende stappen nodig op ieder gebied afzonderlijk. Onderzoek op het gebied van de tekstwetenschap richt zich, onder andere, op het analyseren van teksten en het toekennen van tekststructuur. Maar het was nog de vraag of tekststructuur ook op een betrouwbare manier wordt toegekend. Onderzoek op het gebied van prosodie richt zich, onder andere, op het meten van prosodische kenmerken. De kenmerken die voor tekstprosodie als typisch relevant worden beschouwd zijn pauzeduur, toonhoogte en spreesnelheid. Pauzering en spreesnelheid zijn weinig problematische maten omdat ze direct afleidbaar zijn uit het spraaksignaal. Voor de toonhoogte van een hele zin was het echter nog de vraag of deze op een betrouwbare manier kan worden gekarakteriseerd. In hoofdstuk 2 wordt een onderzoek naar de betrouwbaarheid van tekststructuur-analyses beschreven, in hoofdstuk 3 naar de betrouwbaarheid van toonhoogtemetingen.

In Hoofdstuk 2 wordt de betrouwbaarheid nagegaan van vier procedures om hiërarchische structuur aan een tekst toe te kennen. Bij twee procedures werd gebruik gemaakt van de intuïties van taalgebruikers, en bij twee werd expliciet gebruik gemaakt van theorieën over tekststructuur. De intuïtieve procedures hielden in dat is gevraagd om in een viertal teksten met streepjes aan te geven waar belangrijke grenzen tussen teksteenheden zaten. Deze procedure werd toegepast in een meer en minder beperkende variant: men was vrij om te beslissen hoeveel streepjes men in een tekst zette, óf men kreeg precies te horen hoeveel streepjes men in de tekst mocht zetten. De theorie-georiënteerde procedures hielden in dat aan ervaren tekstwetenschappers werd gevraagd om van ieder van de vier teksten een volledige tekstanalyse te maken. Drie collega-onderzoekers pasten de theorie van Grosz en Sidner (1986) toe met gebruikmaking van de handleiding van Nakatani, Grosz, Ahn en Hirschberg (1995); zes andere collega's pasten de theorie van Mann en Thompson (1986) toe, die bekend staat als Rhetorical Structure Theory. Door deze vier procedures toe te passen op dezelfde vier teksten, is voor iedere tekst een aantal hiërarchische structuren verkregen waarmee de betrouwbaarheid zowel binnen als tussen de theoretische kaders kon worden nagaan. Voor elk van de vier procedures zijn de niveaus van de hiërarchische structuren uitgedrukt in getalsmatige scores. Op grond van deze scores is per procedure de betrouwbaarheid tussen de analisten statistisch geëvalueerd. Van de intuïtieve procedures bleek de beperkende variant het meest betrouwbaar te worden toegepast, en van de theoretisch georiënteerde procedures Rhetorical Structure Theory. Vanwege de hoge betrouwbaarheid en de inhoudelijke specificatie van de relaties tussen teksteenheden is ervoor gekozen om in het verdere onderzoek alleen gebruik te maken van RST.

In hoofdstuk 3 wordt de betrouwbaarheid nagegaan van twee manieren om de toonhoogte van een uiting te meten, als de hoogste piek in de toonhoogtecontour (Lieberman & Pierrehumbert, 1984) óf als de declinatielijnen in de contour ('t Hart et al., 1990). Aan vijf getrainde fonetici werd gevraagd om van veertig zinnen die uit voorgelezen teksten waren gehaald, het F0-maximum en de declinatielijnen (toplijn en basislijn) te bepalen. Ook werden de F0-maxima met behulp van een automatische methode gemeten. De overeenstemming tussen de beoordelaars was hoog met betrekking tot het F0-maximum. De correlaties tussen de F0-maxima bepaald door de vijf beoordelaars en gemeten met de automatische methode waren ook hoog. De overeenstemming tussen de beoordelaars was minder hoog voor de declinatielijnen. De correlaties tussen de F0-maxima en de declinatieparameters waren evenwel hoog, wat erop wijst dat de declinatieparameters voor een belangrijk deel werden gevat door het F0-maximum, in elk geval in de voorgelezen, niet-geëmotioneerde spraak zoals in dit onderzoek gebruikt is. Vanwege de hoge betrouwbaarheid en de hoge correlatie tussen de beoordelaars en de automatische methode, is ervoor gekozen om in het verdere onderzoek naar de relatie tussen tekststructuur en prosodie het F0-maximum te gebruiken als maat voor toonhoogte.

In hoofdstuk 4 wordt de laatste voorbereidende stap voor het onderzoek naar de relatie tussen prosodie en tekststructuur uitgewerkt. Een aantal tekstenkenmerken kan direct afgeleid worden uit de RST analyses die verkregen zijn bij het eerdere onderzoek naar de betrouwbaarheid van tekstanalyses, zoals de syntactische status en de nucleariteit van de teksteenheden, en de

retorische relaties tussen de teksteenheden. Om de teksteenheden, in het vervolg 'segmenten' genoemd, te scoren voor hun niveau in de hiërarchische structuur bestaan echter verschillende mogelijkheden. In hoofdstuk 4 worden drie procedures om de hiërarchische structuur te kwantificeren nader uitgewerkt: van boven naar beneden, van beneden naar boven, en in beide richtingen, een symmetrische procedure. Ook worden twee benaderingen geëxploreerd om de niveaus van segmenten in de hiërarchische structuur te relateren aan de prosodische kenmerken: een relatieve en een absolute. Bij de relatieve benadering worden alleen de prosodische kenmerken van paren van segmenten met elkaar vergeleken. Per paar wordt gekeken welk segment hoger staat in de hiërarchische structuur. In de relatieve benadering van lineaire aangrenzendheid worden alleen paren van segmenten die opeenvolgend zijn in de tekst met elkaar vergeleken; in de relatieve benadering van hiërarchische aangrenzendheid worden alleen paren van segmenten die een dominantierrelatie hebben in de hiërarchische structuur met elkaar vergeleken. Bij de absolute benadering worden de prosodische kenmerken van segmenten van een bepaald niveau in de hiërarchische structuur vergeleken met de prosodische kenmerken van segmenten op alle andere niveaus.

Daarmee is hoofdstuk 4 tevens een eerste verkenning van het onderzoek naar de relatie tussen tekststructuur en prosodie. De teksten waarvan in Hoofdstuk 2 gebruik is gemaakt en waarvan de tekstanalyses beschikbaar waren, waren teksten die op de radio zijn uitgezonden. Het waren twee nieuwsberichten en twee columns die ieder door de auteur werden voorgelezen. Omdat de sprekers de tekst zelf geschreven hadden, kan worden verondersteld dat zij een gedetailleerde mentale representatie hadden van de structuur van tekst en dat zij die prosodisch zouden markeren. Van ieder segment van de tekst werden drie prosodische kenmerken gemeten: de pauzeduur die aan het segment vooraf ging, de toonhoogtepiek en de articulatiesnelheid ervan. Vanwege de grote individuele verschillen tussen sprekers, bijvoorbeeld alleen al omdat vrouwen op hogere toon spreken dan mannen, zijn de prosodische gegevens per spreker gestandaardiseerd. De prosodische kenmerken werden vervolgens gerelateerd aan de niveaus van de segmenten in de hiërarchische structuur. Uit zowel de relatieve als de absolute benadering bleek dat het niveau in de tekststructuur door pauzeduur en toonhoogte wordt gemarkeerd: naarmate grenzen tussen segmenten zich op een lager niveau in de hiërarchische structuur bevinden, zijn de pauzeduren korter en de toonhoogtepieken lager. Er werden geen verschillen gevonden in articulatiesnelheid. De drie procedures om de hiërarchische structuur te kwantificeren lieten hoegenaamd geen verschillen zien in de prosodische realisering. Daarom werd besloten om in het verdere onderzoek naar de relatie tussen tekststructuur en prosodie te werken met de procedure die de relatief duidelijkste resultaten te zien gaf, namelijk de symmetrische procedure. De relatieve en absolute benadering bleven gehandhaafd.

In Hoofdstuk 5 komen de twee gebieden van onderzoek bijeen. De studie die in dit hoofdstuk gerapporteerd wordt, is een corpusstudie naar de prosodische realisering van segmenten in relatie tot hun tekststructurele kenmerken. Er werd gebruik gemaakt van één tekstsoort, namelijk nieuwsberichten. De twintig teksten waren afkomstig uit een landelijk dagblad; de onderwerpen waarover geschreven werd waren divers. De nieuwsberichten hadden een gemiddelde lengte van dertig segmenten. Ze werden geanalyseerd met behulp van Rhetorical Structure Theory. Op basis

van de tekstanalyses werden per segment drie tekststructurele kenmerken vastgesteld: het niveau van het segment in de hiërarchische structuur, de nucleariteit van het segment, en de retorische relatie die het segment met een ander segment onderhield uitgedrukt in termen van causaliteit en semanticaliteit. De teksten werden in doorlopende vorm, dat wil zeggen, zonder typografische middelen waaruit de tekststructuur zou kunnen blijken, aangeboden aan twintig sprekers, één spreker per tekst. De pauzeduur die aan elk segment voorafging werd gemeten in milliseconden, de toonhoogtepiek van ieder segment in hertz, en de articulatiesnelheid als het aantal fonemen dat per seconde werd uitgesproken. De prosodische gegevens werden gestandaardiseerd per spreker. Vervolgens zijn ze in verband gebracht met hun tekststructurele kenmerken. Het niveau in de hiërarchische structuur, nucleariteit, causaliteit en semanticaliteit bleken alle prosodisch gemarkeerd te worden, zij het op verschillende manieren. Voor hiërarchische structuur waren de resultaten vergelijkbaar met die uit Hoofdstuk 4: het niveau in de tekststructuur werd gemarkeerd door pauzeduur en toonhoogte, in die zin dat pauzes langer duren en de toonhoogte hoger is naarmate segmenten op een hoger niveau in de hiërarchie zitten. Nucleariteit bleek gemarkeerd te worden door articulatiesnelheid: segmenten die als nucleus gekarakteriseerd waren, dat wil zeggen, belangrijker waren voor de samenhang in de tekst, werden langzamer voorgelezen dan segmenten die als satelliet gekarakteriseerd waren. Causaliteit werd gemarkeerd door pauzeduur en articulatiesnelheid: de pauzes voorafgaand aan causaal gerelateerde segmenten duurden korter dan de pauzes voorafgaand aan niet-causaal gerelateerde segmenten, en causaal gerelateerde segmenten werden sneller gelezen dan niet-causaal gerelateerde segmenten. Tussen semantisch en pragmatisch gerelateerde segmenten werden geen verschillen in prosodie gevonden.

Het natuurlijk voorkomende tekstmateriaal waarvan in deze studie gebruik gemaakt werd, bevatte allerlei factoren die de resultaten mogelijk beïnvloed hebben. Met name bij de invloeden op de prosodische markerings van causale en niet-causale relaties, en van semantische en pragmatische relaties. De segmenten verschilden namelijk van elkaar qua inhoud en lengte; ze kwamen voor op verschillende plaatsen in de tekst en op verschillende niveaus in de hiërarchische structuur; de tekstrelaties konden lexicaal gemarkeerd en ongemarkeerd voorkomen. Bovendien werden niet alleen de tekstrelaties tussen individuele segmenten in het onderzoek betrokken maar ook die tussen grotere teksteenheden. Tenslotte konden de tekstrelaties in verschillende volgordes voorkomen: bijvoorbeeld causale relaties in oorzaak-gevolg-volgorde en in gevolg-oorzaak-volgorde en pragmatische relaties in feit-commentaar-volgorde en in commentaar-feit-volgorde.

Om specifieke hypothesen over de invloed van tekstrelaties op de prosodische markerings te toetsen zijn twee experimenten uitgevoerd waarover in Hoofdstuk 6 gerapporteerd wordt, een over de invloed van causaliteit op de prosodische realisering en een over de invloed van semanticaliteit. Voor beide experimenten zijn teksten geconstrueerd waarin targetzinnen opgenomen waren. De targetzinnen waren ofwel causaal of niet-causaal verbonden met de voorafgaande zin, ofwel semantisch of pragmatisch. Om de prosodische kenmerken valide met elkaar te kunnen vergelijken moesten de targetzinnen identiek zijn. Dat hield in dat de tekstrelaties niet gemarkeerd konden worden met connectieven en dat geheel uit de context van de targetzin duidelijk moest worden om wat voor relatie het ging, causaal of niet-causaal respectievelijk semantisch of pragmatisch. Op basis van vooronderzoeken zijn uit een

verzameling van geconstrueerde teksten de teksten geselecteerd waarin de tekstrelaties inderdaad duidelijk van elkaar onderscheiden werden. In de eigenlijke experimenten lezen meer dan twintig sprekers de geselecteerde teksten hardop voor. Voor het experiment met betrekking tot de prosodische realisering van causaliteit waren dat zestien teksten met targetzinnen die voorkwamen zowel in een causale als een niet-causale conditie; voor het experiment met betrekking tot de prosodische realisering van semanticaliteit waren dat twaalf teksten met targetzinnen die voorkwamen zowel in een semantische als een pragmatische conditie. Alvorens de tekst hardop voor te lezen lezen de sprekers de tekst verschillende malen voor zichzelf door zodat zij zich bewust zouden worden van de relaties die tussen de zinnen van de tekst bestonden. In het spraakmateriaal werden de pauzeduren voorafgaand en volgend op de targetzin gemeten, de gemiddelde toonhoogte en de toonhoogtepiek van de targetzin, en de articulatiesnelheid van de targetzin. De sprekers bleken de causaal verbonden zinnen met een iets hogere toonhoogte te realiseren dan de niet-causaal verbonden zinnen; en de pragmatisch verbonden zinnen met langere voorafgaande pauzes en een hogere toonhoogte dan semantisch verbonden zinnen.

De resultaten van de experimentele studies weken op een aantal punten af van de resultaten van de eerdere corpusstudie. Met name was dat het geval voor causaal verbonden segmenten. Uit de corpusstudie bleek dat deze vooraf werden gegaan door een kortere pauze en sneller werden gelezen dan niet-causaal verbonden segmenten. Het experiment toonde juist een effect aan op toonhoogte: causaal verbonden segmenten hadden een iets hoger toonhoogtegemiddelde en een iets hogere toonhoogtepiek dan niet-causaal verbonden segmenten. Het feit dat de tekstrelaties in de targetzinnen in het experiment niet lexicaal gemarkeerd waren, kan van invloed zijn geweest voor dit resultaat, in de zin dat prosodie het enig overgebleven middel was waarmee een spreker de tekstrelatie kon aangeven. De bijdrage van prosodie kan in het geval van lexicale ongemarkeerdheid groter zijn dan bij lexicale markerings. In een experimenteel vervolgonderzoek kan deze verklaring worden getoetst. De resultaten voor semantisch en pragmatisch verbonden segmenten uit het experiment waren geheel in overeenstemming met die uit Hoofdstuk 5. Dat pragmatisch verbonden segmenten vooraf worden gegaan door een langere pauze en een hogere toonhoogtepiek hebben dan semantisch verbonden segmenten kan worden beschouwd als het logische gevolg van het feit dat de spreker met een pragmatisch verbonden segment een verschuiving van perspectief aangeeft. Hij of zij doorbreekt hiermee de beschrijving van de gebeurtenissen door bijvoorbeeld een persoonlijke conclusie te trekken of commentaar te geven. De sprekers in het experiment hebben kennelijk geregistreerd dat in de teksten deze verschuiving van perspectief optrad en deze verschuiving vervolgens in hun uitspraak prosodisch gemarkeerd.

De resultaten van de onderzoeken gerapporteerd in de hoofdstukken 4, 5 en 6 kunnen een bijdrage leveren aan de verbetering van tekst-naar-spraak systemen, in de zin dat de prosodische parameters in deze systemen aangepast kunnen worden op de manier waarop menselijke sprekers deze tekststructurele kenmerken prosodisch realiseren. Wanneer de tekststructuur van teksten bekend is, kunnen de hiërarchische niveaus aangegeven worden met de systematische variatie in pauzeduur en toonhoogte zoals uit de studies naar voren is gekomen. Dat geldt ook voor semantische en pragmatische relaties, omdat de resultaten daarvan eenduidig zijn en goed

Samenvatting

geïnterpreteerd kunnen worden. Voor causale en niet-causale relaties is dat nog niet in voldoende mate het geval.

De resultaten van deze studies zijn ook relevant voor de theorievorming over tekstproductie. Wanneer de planningsfactor bij het spreken is uitgeschakeld zoals in deze studies is gebeurd door gebruik te maken van voorbereide voorgelezen spraak, en sprekers in staat blijken om tekststructurele kenmerken zoals hiërarchie, nucleariteit, causaliteit en semanticaliteit, prosodisch te markeren, dan toont dit aan dat deze kenmerken psychologisch relevant zijn: de reflectie ervan in prosodie laat zien dat zij onderdeel uitmaken van de mentale representatie van de tekst.

Curriculum vitae

Hanny den Ouden was born in Hendrik Ido Ambacht, the Netherlands, on 22 March 1964. From 1976 to 1982, she followed pre-university education at Juvenaat H. Hart in Bergen op Zoom (Gymnasium alpha). She started to study Theology at Tilburg University. After the propaedeutic exam in 1983, she exchanged university for a vocational training in psychiatric nursing. She gained her nursing qualification in 1987. She then returned to the study of Theology and sat her candidate exam in 1991, working at the same time as a nurse in the psychiatric clinic 'Het Hooghuys' in Etten-Leur.

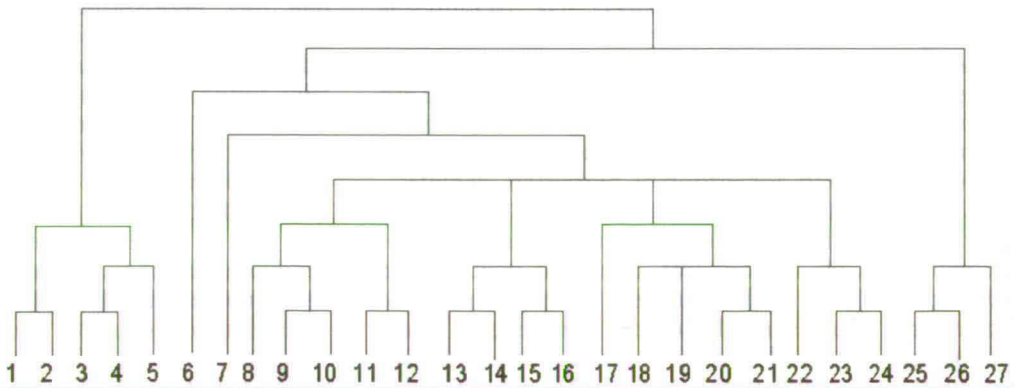
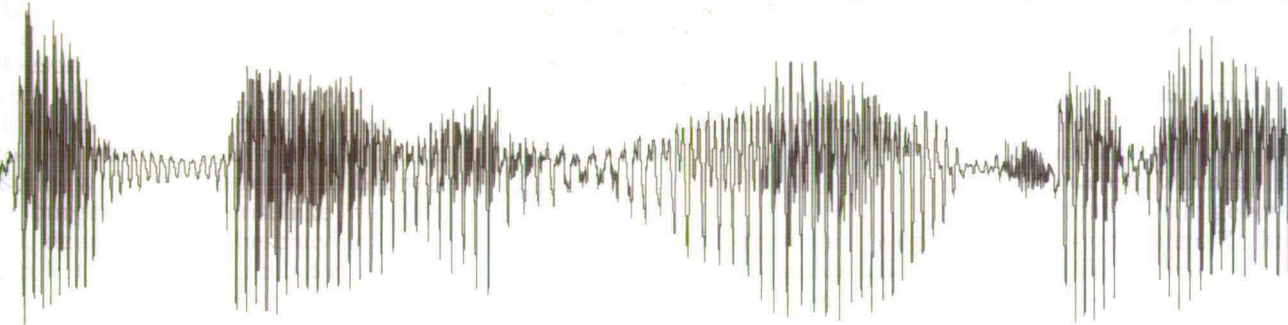
From 1989 to 1995, she studied Linguistics at Tilburg University. She graduated (cum laude) in Discourse Studies. Her three children were born in the years 1991, 1993, and 1996. From 1997 to 2002, she was employed as a Ph.D. student at IPO, Technische Universiteit Eindhoven, and at the Discourse Studies Group, Tilburg University. In that period, she worked on the studies reported in the present thesis.

From 2002 to 2004, she worked as a lecturer of Methodology and Statistics at the Vrije Universiteit Amsterdam. Since September 2004, she holds a position as assistant professor of Language use and Discourse studies in the department of Dutch Language and Culture and the Utrecht Institute of Linguistics OTS at Utrecht University, The Netherlands.

Bibliotheek K. U. Brabant



17 000 01543222 3



J.F. Schouten School for
User-System Interaction Research